1

Introduzione

OBIETTIVI

- Delineare il contesto in cui si situano gli argomenti trattati in questo libro
- Tratteggiare l'evoluzione dei calcolatori elettronici, ricostruendone la storia dagli albori
- Richiamare i nomi dei pionieri che con le loro ricerche, studi e realizzazioni hanno aperto campo di attività che ha finito per interessare ogni aspetto del mondo contemporaneo
- Tratteggiare la produzione dei primi calcolatori commerciali e quella dei microprocessori
- Identificare le classi di computer del tempo presente
- Illustrare il tasso di sviluppo quantitativo dell'elettronica digitale e commentare le implicazioni che esso comporta
- Definire i concetti di "Architettura" e "Organizzazione"

CONCETTI CHIAVE

Calcolatore elettronico numerico, unità di elaborazione, memoria, microprocessore, classi di sistemi, evoluzione tecnologica, legge di Moore, consumi, livelli di astrazione, architettura, organizzazione.

INTRODUZIONE

Questo capitolo cerca di costruire il contesto entro il quale si collocano gli argomenti affrontati dal libro. Si inizia con una narrazione dei passi fondamentali dell'evoluzione dei calcolatori elettronici, richiamando i nomi dei matematici, degli scienziati, degli ingegneri che di tali passi sono stati gli artefici.

Successivamente l'attenzione è rivolta alle macchine di produzione industriale. In particolare si fa un resoconto di come sono nati e come si sono sviluppati i microprocessori, dal primo micro a 4 bit agli attuali dispositivi 64 bit. Viene quindi proposta una classificazione dei sistemi correnti in base al tipo di impiego e di campo applicativo. Riferendoci alla cosiddetta "legge di Moore", viene svolto un esame circa la crescita delle prestazioni dei dispositivi elettronici, con particolare riferimento alla crescita delle prestazioni dei tre principali componenti dei sistemi di elaborazione: processori, memorie e memorie a disco. In chiusura, viene definito il significato di alcuni termini ricorrenti, come "architettura", "organizzazione" e "modello di programmazione".

Nell'esposizione è spesso necessario fare ricorso a termini e a concetti non ancora definiti. Si fa leva sull'intuito del lettore.

1.1 Qualche cenno storico

Una storia dell'evoluzione dei calcolatori elettronici costituirebbe da sola materiale sufficiente per la scrittura di un volume di carta stampata. Qui di seguito richiameremo solo alcuni esempi di macchine che, per l'epoca in cui sono state concepite e per i concetti in esse incorporati, sono da considerarsi pietre miliari.

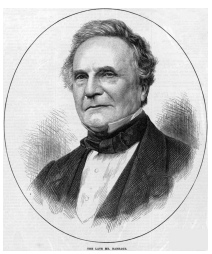
Si usa suddividere l'evoluzione dei calcolatori elettronici in almeno 4 generazioni, in dipendenza dalla tecnologia dell'epoca e in corrispondenza agli archi di tempo (orientativi) elencati.

- Prima generazione, 1940 - 1950, nella quale erano impiegate valvole termoioniche.
- Seconda generazione, 1950 - 1964, nella quale erano impiegati transitori.
- Terza generazione, 1964 - 1971, nella quale si faceva uso di circuiti integrati.
- Quarta generazione, 1971 - tempo presente, dove si fa uso di microprocessori.

Alle generazioni appena elencate se ne potrebbe aggiungere una quinta. Ci riferiamo alle unità di elaborazione che fanno parte di tablet e telefoni portatili, o che costituiscono il cuore di un numero potenzialmente infinito di apparati di uso quotidiano (orologi, macchine fotografiche, elettrodomestici, e via elencando fino ai *wearable computer*). Caratteristiche di spicco di questa generazione sono, oltre all'alto livello di integrazione, le dimensioni ridotte e i bassissimi consumi.

1.1.1 Le tappe fondamentali

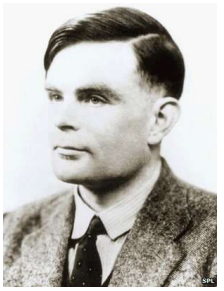
Nel 1837, Charles Babbage, un professore di matematica dell'Università di Cambridge, Gran Bretagna, propose una macchina, denominata *Analytical Engine*, che, pur essendo meccanica, era concepita in accordo ai concetti che, più di secolo dopo, avrebbero dato vita ai calcolatori elettronici. Qualche anno prima, Babbage aveva progettato una macchina, denominata *Difference Engine*, il cui uso era limitato al solo calcolo di tavole matematiche. La Difference Engine, come le macchine calcolatrici di qualche secolo prima, la "Pascalina" costruita da Pascal e la "Machina aritmetica" di Leibnitz, poteva effettuare solo il genere di calcolo per il quale era stata pensata. Mentre stava costruendo la Difference Engine, Babbage si rese conto che sarebbe stato possibile realizzare una macchina di maggior potenzialità. Essa avrebbe dovuto essere guidata da un *programma*, in grado di comandare sequenze di operazioni. In pratica, Babbage delineò il primo *calcolatore di uso generale*. L'Analytical Engine aveva una memoria, detta magazzino (*store*) e una unità aritmetica detta mulino (*mill*); leggeva/scriveva dati in ingresso/uscita su schede perforate. Il magazzino era composto da registri di 50 cifre (digit) decimali; ciascun registro era costituito da una pila di ruote dentate montate sullo stesso asse, ciascuna delle quali rappresentava un digit. L'angolo di cui ciascuna ruota era sfasata rispetto alla posizione di riferimento determinava il valore del digit corrispondente: per la lettura, la ruota veniva riportata alla posizione di riferimento, la misura di quanto doveva essere ruotata corrispondeva al valore del digit. Il magazzino era usato per contenere dati e risultati dell'elaborazione. Il mulino prelevava i dati dal magazzino, effettuava operazioni aritmetiche su di essi ed eventualmente memorizzava i risultati nel magazzino stesso. Un complesso meccanismo fatto di pignoni, barre dentate e leve, permetteva di azionare i registri del magazzino. Il programma era pure su schede perforate. La macchina era in grado di eseguire le istruzioni contenute in forma codificata sulle schede perforate. Tra le istruzioni erano



Charles Babbage

previsti anche i salti condizionati, in modo da consentire flussi alternativi nell'esecuzione del programma. Data l'epoca, l'Analytical Engine era completamente meccanica; il suo azionamento avrebbe dovuto essere a vapore.
A parte il fatto di non usare l'aritmetica binaria e di non avere il programma memorizzato, l'Analytical Engine incorporava sostanzialmente tutti i concetti che stanno alla base di quelli che oggi chiamiamo calcolatori elettronici. Purtroppo non ne venne mai realizzato un prototipo funzionante.

Un secolo dopo, nel 1936, Alan Matison Turing pubblicava un articolo, intitolato *On computable numbers, with an application to the Entscheidungsproblem* nel quale veniva definito un modello di macchina, poi denominato "Macchina di Turing" (MdT), che da allora costituisce il modello formale di macchina universale di calcolo. Il "Problema della decisione" (*Entscheidungsproblem*) era stato posto nel 1928 dal matematico tedesco David Hilbert. Detto in maniera semplificata, esso richiedeva di esibire una procedura in grado di stabilire, per un qualsiasi enunciato matematico, se esso fosse vero o falso, in un numero finito di passi. Turing propose un modello di macchina, composto da un nastro potenzialmente infinito, diviso in celle, e da una testina di lettura/scrittura, in grado di spostarsi avanti e indietro lungo il nastro e di leggere/scrivere simboli dalle/nelle celle sul nastro stesso. La macchina era stata concepita secondo il criterio di simulare i processi di calcolo degli esseri umani: una successione di passaggi nei quali si applicano regole definite con la produzione, ad ogni passaggio, di risultati intermedi registrati su carta e ulteriormente elaborati ai passaggi successivi.


Alan Turing

Non ci soffermeremo a descrivere la struttura e il funzionamento della macchina di Turing, rimandando il lettore all'immensa bibliografia che esiste sull'argomento. Diremo soltanto che essa formalizza i concetti di *algoritmo*, *computazione*, *calcolatore programmabile* e *programma memorizzato*. Ed è in grado di eseguire tutte le elaborazioni effettuabili mediante le macchine costruite dall'uomo[1]. Tuttavia, si tiene a precisare che la Macchina di Turing è un modello teorico, non un modello di architettura dei calcolatori. Invariabilmente chi costruisce esemplari che ne rispecchiano il funzionamento lo fa a puro scopo dimostrativo o didattico. Chi fosse interessato alla figura di Turing faccia riferimento a [Hod14], chi volesse approfondire lo sviluppo del pensiero che ha portato da Leibniz a Turing ed oltre può fare riferimento a [Dav12]e [Dys12].

Nel 1936, un ingegnere tedesco di nome Konrad Zuse iniziò la costruzione di un calcolatore programmabile di uso generale, denominato "Z1", e realizzato con relé elettromeccanici. Questo prototipo usava la rappresentazione binaria ed era in grado di effettuare operazioni in virgola mobile. La memoria era meccanica. Le istruzioni (il programma) venivano lette da un nastro di celluloide perforato, simile ad una pellicola cinematografica, sul quale venivano poi scritti anche i risultati dell'elaborazione. La macchina venne completata nel 1938. Ma era continuamente soggetta a guasti a causa delle parti meccaniche. Concettualmente, rispetto a come si sarebbero sviluppati i calcolatori elettronici, la macchina era

Konrad Zuse

[1]Turing è oggi ritenuto dalla comunità scientifica come uno dei massimi matematici/logici del secolo scorso. Il contributo che egli ha dato con la macchina che porta il suo nome e quello dato con l'articolo intitolato *Computing machinery and intelligence* (1950), con il quale pose le basi dell'intelligenza artificiale, fanno considerare Turing come il padre della "scienza informatica".

deficitaria per un solo aspetto: non prevedeva i salti condizionati. Al di là di questo, il prototipo Z1 è stato il primo calcolatore programmabile funzionante, anche se pochissimo affidabile. Nel 1943, i bombardamenti su Berlino distrussero il prototipo Z1 e tutti i relativi documenti di progetto. Dopo la guerra lo stesso Zuse ricostruì un prototipo di Z1 che ora si trova al museo della tecnica di Berlino.

Nel 1939, John Atanasoff, professore di matematica e fisica alla Iowa University, assieme allo studente di ingegneria Clifford Berry, costruì un prototipo di calcolatore detto ABC (*Atanasoff-Berry Computer*) il cui scopo era facilitare il calcolo di sistemi di equazioni lineari. Questa macchina non era un calcolatore di uso generale, non essendo programmabile, ma introduceva alcune innovazioni che sarebbero diventate standard negli anni a seguire e che hanno indotto più parti a considerare ABC come il primo calcolatore elettronico. Le innovazioni più appariscenti erano l'uso del sistema binario e l'impiego esclusivo di circuiti elettronici per il calcolo, senza l'intervento di parti meccaniche o relé elettromeccanici. Inoltre la sua memoria era composta da condensatori i cui contenuti andavano rigenerati, una tecnica concettualmente simile a quella delle odierne memorie DRAM.

John Atanasoff

Ma il primo calcolatore elettronico programmabile non entrò in funzione che nel 1944. Si trattava del *Colossus*, messo in opera nel Regno Unito, durante la seconda guerra mondiale. Impiegava oltre 2.000 valvole termoioniche. Venne costruito in segreto per la Royal Navy a Bletchley Park, il luogo a nord di Londra, dove, sotto l'egida del governo inglese, era stato radunato un folto gruppo di scienziati ed esperti, con l'obiettivo di decifrare le comunicazioni dei tedeschi durante la seconda guerra mondiale. Colossus venne delineato da Max Newman, un matematico inglese, ispirandosi ai concetti della macchina di Turing[2]. La realizzazione pratica di Colossus fu comunque opera dell'ingegnere britannico Tommy Flowers, cui si deve il merito di essere riuscito a convincere i membri del gruppo ad impiegare le valvole termoioniche, allora non ben viste rispetto ai relé convenzionali. Colossus veniva tenuto sempre acceso, visto che le valvole tendevano a guastarsi in seguito ad accensioni/spegnimenti. Colossus non era a programma memorizzato: la programmazione si effettuava attraverso spinotti e interruttori. Per ordine di Winston Churchill, a fine della guerra la macchina fu distrutta e i progetti relativi bruciati. Solo negli anni novanta, quando ormai le ragioni di segretezza erano del tutto tramontate, venne resa nota la sua realizzazione.

Tommy Flowers

La segretezza che ha circondato il Colossus ha fatto ritenere per molto tempo che il primo calcolatore elettronico sia stato progettato e costruito da J. Presper Eckert e John Mauchly, dell'Università della Pennsylvania[3]. La macchina venne denominata ENIAC per *Electronic Numerical Integrator and Calculator* e la sua costruzione venne finanziata dal Ministero della Difesa USA durante la Seconda guerra mondiale. La macchina veniva impiegata per calcoli balistici. La sua esistenza venne resa nota nel 1946, a

[2]Turing faceva parte del gruppo di scienziati di Bletchley Park ed era impegnato a decifrare i codici di *Enigma*. Newman invece lavorava a decifrare le comunicazioni via telescrivente che i tedeschi cifravano con un apparato, prodotto dalla ditta tedesca Lorenz, diverso da Enigma.

[3]Vale la pena di ricordare che Eckert e Mauchly, come pure Zuse e Atanasoff non avevano conoscenza della macchina di Turing.

guerra finita. Occupava un volume immenso: aveva forma di U, era lunga circa 30 metri, alta 2 e spessa un metro, per oltre 120 metri cubi di volume, su una superficie di 180 m^2. Pesava 30 tonnellate. Impiegava oltre 18.000 tubi a vuoto (valvole) collegati da 500.000 contatti saldati a mano. Aveva 20 registri di 10 cifre decimali; ogni registro era lungo oltre mezzo metro. Consumava poco meno di 180 kW e si racconta che un intero quartiere di Filadelfia andò al buio quando venne messa sottotensione per la prima volta. ENIAC era in grado di eseguire 1900 somme al secondo, ovvero circa 10,5 somme/kW. La programmazione avveniva in modo manuale, attraverso fili e interruttori sul pannello di controllo, non dissimile da quello di una vecchia centralina telefonica. I dati venivano introdotti attraverso schede perforate. Le schede perforate saranno per anni lo standard di ingresso, non solo per i dati, ma anche per i programmi[4]. Mentre Atanasoff non aveva brevettato il suo ABC, Mauchly ed Eckert brevettarono l'ENIAC. Ma, nel 1973, un giudice federale americano ha annullato il brevetto di Mauchly ed Eckert, stabilendo che l'ENIAC derivava dal computer di J. Atanasoff e C. Berry, costruito nel 1939. Al di là delle opinioni del giudice, resta il fatto che quello di Atanasoff non era un computer programmabile.

Mauchly e Eckert

Il problema principale di ENIAC era la modalità di programmazione. Nel Giugno del 1945, John Von Neumann, che faceva parte sin dal 1944 del gruppo di ricercatori e progettisti che ruotava attorno al Ministero della Difesa USA ed era entrato in contatto con Mauchly ed Eckert per il progetto ENIAC, scrisse un rapporto intitolato *First Draft of a Report on the EDVAC* in cui veniva esposta per la prima volta l'idea di costruire un calcolatore a programma memorizzato, che egli chiamò EDVAC (*Electronic Discrete Variable Automatic Computer*). Von Neumann era di origine ungherese, naturalizzato americano e professore di matematica a Princeton. All'epoca era considerato il più importante matematico vivente[5] e collaborava con il Progetto Manhattan per la realizzazione della bomba atomica. Nel 1937 aveva conosciuto Turing a Princeton, dove questi aveva vinto una borsa di studio, con il sostegno dello stesso von Neumann. La soluzione proposta da Von Neumann diventò la norma e da allora si iniziò a parlare di "Modello di Von Neumann", ovvero di "Architettura di Von Neumann"[6]. L'architettura di Von Neumann, su cui torneremo nel seguito, è schematizzata in Figura 1.1. Sostanzialmente si tratta dell'organizzazione comune a tutti quelli che oggi chiamiamo calcolatori elettronici, a parte il fatto che Unità di controllo e Unità aritmetico-logica sono oggi raggruppati in un unico componente denominato CPU (*Central Processing Unit*). La caratteristica fondamentale è quella di avere una sola memoria, nella quale vengono contenuti dati e programmi.

John Von Neumnn

[4]Fino alla metà degli anni settanta del secolo scorso, perforatori e lettori di schede erano parte del normale paesaggio di qualunque centro di calcolo.

[5]Di lui si racconta che fosse un genio multiforme. Si dice che a dieci anni padroneggiasse 10 lingue e fosse capace di effettuare a mente divisioni di 8 cifre. Era anche un novello Pico della Mirandola, essendo capace di ripetere a memoria intere pagine di libri letti anni prima.

[6]Hennessy e Patterson [HP06] espongono una serie di argomenti in base ai quali la fama di Von Neumann è da considerarsi parzialmente usurpata, concludendo che la gloria spetterebbe a Eckert e Mauchly per essere stati i primi a realizzare un calcolatore elettronico funzionante. Affermazione che, dalla precedente narrazione, non risulta vera.

Istruzioni e dati sono codificati in forma binaria e, di per sé, non sono distinguibili le une dagli altri. La sequezialità del programma è ottenuta attraverso uno speciale registro della CPU, il *Program Counter* ("contatore di programma") che indica il punto cui è pervenuta l'esecuzione del programma stesso. Nonostante la memoria finita, il modello di Von Neumann ha la potenzialità della Macchina di Turing [Neu14].

Con Von Neumann si conclude il racconto sui personaggi che hanno dato contributi fondamentali allo sviluppo dei calcolatori. Il lettore interessato potrà trovare in Internet una vasta schiera di scienziati, ingegneri, inventori che pure hanno dato importanti contributi e che, per brevità abbiamo omesso di menzionare.

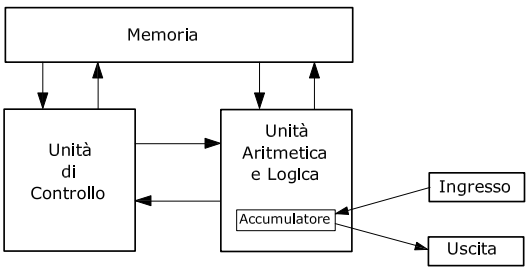


Figura 1.1 Il modello originale di calcolatore elettronico di Von Neumann.

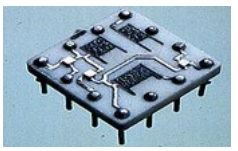
1.1.2 I primi calcolatori commerciali

Eckert e Mauchly fondarono una società che costruì il prototipo chiamato BINAC. La società venne poi acquistata dalla Remington-Rand, per la quale, nel 1951, venne prodotto il primo esempio di calcolatore commerciale: l'Univac I, venduto per circa 250.000 dollari USA. Di Univac I ne vennero prodotti 48 esemplari.

L'IBM, presente da anni nel settore delle macchine da ufficio, produsse il primo suo calcolatore (l'IBM 701) nel 1952, vendendone una ventina. Le vendite furono ritenute un grande successo, soprattutto se si considera che non molti anni prima il suo amministratore delegato, T. J. Watson, aveva profetizzato che sul mercato mondiale "c'era posto per 5 computer". Il 701 impiegava tubi termoionici. Visto il successo del 701, IBM dette impulso alla sua produzione di computer. Seguirono altri modelli di macchine e l'IBM divenne il maggior attore di mercato. Dall'impiego delle valvole si passò ai transistori. Il primo calcolatore IBM che impiegava solo transistori apparve nel 1958.

Nell'Aprile del 1964 IBM annunciò il sistema S/360, come risultato di un progetto di sviluppo nel quale la società aveva investito svariati milioni di dollari. Per la prima volta si distingueva il concetto di architettura da quello di organizzazione o, meglio, di realizzazione, nel senso che l'IBM aveva definito un modello di macchina che sarebbe stato comune a tutta una famiglia di calcolatori. Questi avrebbero potuto avere dimensioni, capacità e potenzialità diverse, avrebbero potuto essere realizzati con tecnologie differenti, ma sarebbero stati comunque compatibili tra di loro. Diversamente dal passato, nello sviluppare una nuova macchina, i progettisti non avrebbero cominciato da zero. Si trattò di un passaggio fondamentale: da quel momento i concetti di architettura e compatibilità avrebbero costituito il fondamento per ogni produzione industriale. Con

il sistema S/360 l'IBM sviluppò la cosiddetta *Solid Logic Technology* (SLT), si trattava di piccoli moduli quadrati (meno di 2 cm di lato), sui quali venivano montati uno o più transistori e diodi discreti, successivamente incapsulati. Non erano ancora dei veri circuiti integrati – che, pure, facevano la comparsa proprio in quegli anni[7] – ma consentivano un consistente risparmio di spazio rispetto all'uso di transistori e diodi separati. La memoria dei calcolatori della famiglia S/360 era a nuclei magnetici: ciascun bit era rappresentato da un minuscolo toroide. Una tecnologia che resterà ampiamente in uso fino metà anni settanta.


Mudulo SLT

L'IBM dominerà per lungo tempo il mercato dei calcolatori in tutto il mondo. Agli inizi degli anni settanta del secolo scorso, essa era padrona di oltre il 60% del mercato, mentre l'insieme dei suoi concorrenti si divideva il restante 40%. A fine anni sessanta, venne introdotto il sistema S/370, [IBM74] successore del precedente, ma sempre compatibile. Da allora, l'IBM non ha mai smesso di costruire macchine sostanzialmente discendenti da quell'architettura, anche se ad asse sono state affiancate linee di prodotto basate su architetture di più larga diffusione. Non ci si deve stupire di tale longevità, in quanto essa è stata determinata dalla necessità di dover salvaguardare i giganteschi investimenti nel software fatti da chi impiegava quelle macchine, largamente diffuse in banche, ministeri o grandi imprese. La necessità di salvaguardare il patrimonio software preesistente è stata, e continua a essere, una delle forze condizionanti l'evoluzione dell'intero mercato dell'informatica.

Gli anni sessanta e settanta sono stati gli anni delle macchine della categoria chiamata *mainframe*. Si trattava di macchine costose (di norma sopra al centinaio di milioni di lire, di allora) equipaggiate con sistemi operativi che ne permettevano un uso condiviso da parte di più utenti. Nel 1963 apparve il CDC 6600, il primo vero supercalcolatore[8]. Il suo progettista Seymour Cray, dopo aver progettato altri calcolatori per la Control Data Corporation, fondò successivamente una sua società che continua a produrre macchine ultrapotenti.

Nel 1965 la DEC (Digital Equipment Corporation), che nel 1960 aveva introdotto il suo primo minicalcolatore (il PDP-1), introdusse il PDP-8, primo minicalcolatore a costare sotto i 20.000 dollari. Il PDP8 aveva un solo registro accumulatore di soli 8 bit.

Con i minicalcolatori l'impiego dei sistemi di elaborazione si diffuse anche entro organizzazioni non dotate di bilanci miliardari. Nacquero nuove industrie per produrre macchine di questa categoria; come pure ci furono industrie attive in altri campi che si misero a fabbricare calcolatori (per esempio, la Hewlett-Packard). Sempre la Digital introdusse nei primi anni settanta il PDP 11. Si trattava di un minicalcolatore a 16 bit organizzato attorno a un unico bus. Un'organizzazione che sarà lo standard per una ventina di anni, specialmente con l'avvento dei microprocessori. Il PDP-11 fu il primo

[7] Il primo circuito integrato venne sviluppato da Jack Kilby della Texas Instruments nel 1958. Robert Noyce, uno dei futuri fondatori di Intel, egli pure impegnato a produrre il primo circuito integrato, arrivò pochi mesi dopo. Per questa invenzione Kilby ricevette il premio Nobel per la fisica nel 2000.

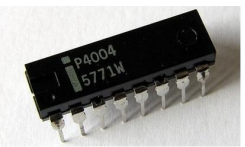
[8] Un CDC 6600, venne installato nel 1969 presso la sede di Bologna del CINECA. Il CINECA venne istituito nel 1969 come Consorzio Interuniversitario per il Calcolo Automatico dell'Italia Nord Orientale. Oggi, dopo essere stato ridenominato più semplicemente "Consorzio Interuniversitario", è il maggiore centro di calcolo in Italia. Opera sotto il controllo del Ministero dell'Istruzione dell'Università e della Ricerca (MIUR). Di esso fanno parte 70 università italiane, 4 Enti di Ricerca Nazionali oltre al MIUR medesimo. L'attività distintiva è fornire supercalcolo alla comunità scientifica. I sistemi del Cineca compaiono regolarmente nella classifica dei TOP 500 (`http://www.top500.org/`); nella classifica di Giugno 2012, la settima posizione era tenuta dal sistema "Fermi", ovvero un IBM BlueGene/Q, installato presso il Consorzio.

sistema a operare sotto lo UNIX sviluppato presso AT&T (scritto in C). E sul PDP-11 fu sviluppato il BSD (*Berkeley Standard Distribution*). Più tardi la DEC introdusse il Vax (definito supermini), estendendo in molti versi l'architettura PDP 11. Il Vax sarà una specie di riferimento per le macchine di categoria intermedia e avrà grande diffusione nel mondo scientifico, accademico e industriale, tanto che per anni è stato costume confrontare con il Vax le prestazioni di nuove macchine immesse sul mercato. A fine anni settanta la DEC occupava saldamente la seconda posizione nella classifica dei produttori di sistemi di elaborazione[9], sebbene assai distanziata quanto a fatturati dal gigante IBM.

Ma ormai era iniziata l'epoca dei microprocessori. Essi avrebbero scombussolato il mondo dei calcolatori, facendo nascere nuovi protagonisti e ridimensionando affermate posizioni di mercato.

1.1.3 I microprocessori

Il termine *microprocessore* venne usato per la prima volta dalla Intel quando introdusse, esattamente il 15 Novembre 1971, il sistema MCS–4, un gruppo di 4 circuiti integrati comprendente una CPU, una memoria ROM, una memoria RAM e un dispositivo per l'espansione delle porte di ingresso/uscita. I quattro integrati consentivano la facile realizzazione di un *computer microprogrammmabile* a 4 bit. La CPU, denominata 4004, conteneva un sommatore parallelo a 4 bit, 16 registri a 4 bit, un (solo) *registro accumulatore* e uno *stack*.



L'integrato 4004.

Il 4004 conteneva al suo interno 2.300 transistori e, al primo rilascio, operava a una frequenza di 108 KHz. Era prodotto in forma di integrato a 16 piedini su due file[10].

I microprocessori hanno rappresentato il punto di svolta nell'uso del calcolatore: da macchina gigantesca, utilizzata solo da pochi addetti ai lavori, il calcolatore elettronico divenne un apparato di dimensioni contenute, convenientemente utilizzato nell'industria, negli uffici e anche a livello domestico.

Alla fine degli anni sessanta si era evidenziato un serio problema nella progettazione dei sistemi elettronici digitali: la loro complessità si stava avvicinando al punto in cui il numero di componenti necessari alla realizzazione delle funzionalità richieste dagli apparati avrebbe superato i limiti del trattabile. Un rimedio consistette nell'utilizzare logica convenzionale accoppiandola a memorie ROM, in funzione di reti combinatorie complesse. Questa soluzione condusse alla nascita del microprocessore. Nell'estate del 1969, il costruttore giapponese Busicom chiese alla Intel di produrre un integrato per una famiglia di calcolatrici programmabili ad alte prestazioni. Il committente intendeva realizzare calcolatrici con una mezza dozzina di integrati, per ciascuno dei quali si prevedevano dai 600 ai 1000 transistori. All'Intel (allora una piccolissima, per quanto brillante società) esaminarono il progetto e conclusero che lo sviluppo di un tale apparato sarebbe stato

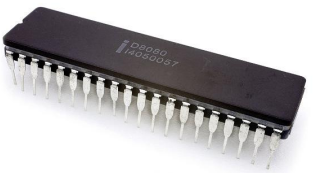
[9]Una decina d'anni dopo l'avvento dei microprocessori, la DEC entrerà in una fase di lento declino e finirà per essere assorbita nel 1998 dalla Compaq, un'industria nata negli anni ottanta, fattasi ricca fabbricando personal computer. Agli inizi del 2002, la Compaq, a sua volta, è stata assorbita dalla Hewlett-Packard.

[10]In lingua inglese, questa organizzazione viene chiamata *Dual In-line Package* (DIP). I microprocessori continuarono a essere prodotti secondo lo schema DIP per circa una decina d'anni dalla data di introduzione del 4004. Successivamente, la necessità di disporre di un sempre maggior numero di piedini fece abbandonare questa organizzazione per altre forme di piedinatura.

troppo complesso per il prezzo di mercato previsto. Alcuni progettisti Intel avevano esperienza d'uso del minicalcolatore PDP-8, al cui interno c'era una ROM piuttosto grande, e pensarono che la tecnologia fosse matura per sviluppare in forma integrata una unità di elaborazione capace di interpretare sequenze di istruzioni generali, contenute in ROM. Un simile processore avrebbe trovato impiego non solo nelle calcolatrici ma anche in luogo dei minicomputer[11]. Nel Novembre del 1971, in cambio di una riduzione del prezzo praticato nei confronti del committente Busicom, per cui era stato sviluppato il progetto, Intel si accordò per la libera distribuzione sul mercato dell'MCS-4. La pubblicazione del primo articolo sul dispositivo suscitò un notevole interesse, a seguito del quale Intel realizzò un modesto ma promettente guadagno di 8.500 dollari.

Mentre stava lavorando al 4004, l'Intel aveva avviato un progetto parallelo che avrebbe portato al primo microprocessore a 8 bit, l'8008. L'8008 fu introdotto nell'Aprile del 1972, nella forma di un integrato a 18 piedini, ed era realizzato in tecnologia PMOS. Integrava 3.500 transistori e aveva un tempo medio di esecuzione delle istruzioni di 30 μs.

Fu però solo con l'introduzione dell'8080 [Noy81] nell'Aprile del 1974, che il microprocessore venne accreditato dalla comunità scientifica e tecnologica come un dispositivo non solo capace di sostituire la logica convenzionale, ma anche in grado di essere impiegato come un elaboratore di uso generale. L'8080 venne prodotto in tecnologia NMOS, per un totale di 4.500 transistori in un integrato a 40 piedini. La frequenza del *clock* poteva raggiungere i 2 MHz consentendo l'esecuzione di alcune istruzioni in 2 μs. l'8080 indirizzava uno spazio di memoria di 64 KB, aveva al suo interno 7 registri a 8 bit, 6 dei quali potevano essere impiegati in coppia a formare registri di 16 bit. Si trattava di un dispositivo con prestazioni tali da essere in breve tempo riconosciuto come la macchina a 8 bit per eccellenza.



L'integrato 8080.

La rapida accettazione da parte del mercato dei microprocessori a 8 bit e le forti richieste di 8080 dettero la spinta a due nuovi microprocessori concorrenti, il Motorola MC6800 e lo Zilog Z80[12]. Il 6800, introdotto a metà del 1974, fu il primo microprocessore ad avere un'alimentazione singola di +5V (la risposta dell'Intel fu l'8085, evoluzione tecnologica – e minimamente architetturale – dell'8080). L'eliminazione delle tensioni di alimentazione multiple abbassò il costo del prodotto e rese il 6800 molto diffuso sul mercato. Lo Z80, realizzato nel 1975, rifletteva i progressi nell'architettura fatti nei due anni successivi all'uscita dell'8080 e del 6800. Lo Z80 è stato considerato come il punto più alto raggiunto dall'evoluzione delle architetture di questa generazione. Incorporava il repertorio di istruzioni dell'8080 e quindi aveva piena compatibilità con i programmi che erano stati scritti per questo.

Il primo microprocessore su singolo integrato con bus dati esterno a 16 bit fu il PACE, un dispositivo MOS, contenuto in un package DIP da 40 piedini, introdotto nel 1974 dalla National. Altri costruttori produssero i loro microprocessori a 16 bit; tra essi la Texas

[11] Un grosso contributo alla creazione dell'MCS-4 fu dato dall'italiano Federico Faggin, in seguito fondatore e presidente della Zilog, il quale, approdato alla Intel nel 1970, afferrò rapidamente il concetto del nuovo dispositivo che si stava sviluppando e si applicò alla sua ottimizzazione in tecnologia MOS: in soli 9 mesi produsse i campioni dei 4 integrati che sarebbero poi diventati l'MCS-4. Si vedano [FHMS96] e [Fag09].

[12] Naturalmente a Silicon Valley e dintorni non c'era solo l'Intel a studiare queste nuove tecnologie, praticamente tutti i produttori di dispositivi elettronici erano impiegati nello sviluppo dei loro microprocessori, cercando di arrivare sul mercato con il massimo anticipo possibile rispetto alla concorrenza.

Instruments col suo TMS9900. Purtroppo questi dispositivi continuavano a indirizzare "solo" 64 KB, perché, all'epoca, questa misura era generalmente giudicata al di sopra di qualunque necessità pratica.

Fu il microprocessore 8086 della Intel, introdotto nel Giugno 1978, con il suo Mega byte di memoria indirizzabile, a determinare il passaggio alla seconda generazione. L'8086 venne prodotto in tecnologia HMOS da 3 μm, per un totale di 29.000 transistori, in un integrato a 40 piedini. La frequenza del clock era di 5 MHz, per una capacità elaborativa media di 0,33 MIPS (Milioni di Istruzioni Per Secondo). L'8086 definì le basi di una nuova architettura, normalmente designata come "architettura ×86". Questa architettura ha costituito la pietra miliare per lo sviluppo dei microprocessori negli anni a seguire, ha condizionato e continua a condizionare lo sviluppo della stessa informatica. Dato il predominio dell'architettura ×86, svariati produttori di dispositivi elettronici, tra cui la Advanced Micro Devices (AMD) seguirono l'Intel nel produrre CPU compatibili[13]. Col passare degli anni questa architettura si è evoluta fino agli attuali processori a 64 bit: la stragrande maggioranza dei PC (*Personal Computer*, ovvero calcolatori personali) sparsi per il mondo ha al suo interno una CPU le cui radici affondano nell'8086.

All'Intel 8086 seguirono presto le versioni a 16 bit dei micro di Motorola e Zilog. La prima introdusse il dispositivo MC68000, la seconda introdusse lo Z8000. In comune all'8086 avevano il bus dati esterno di 16 bit e la possibilità di indirizzare un spazio di memoria allora ritenuto immenso. Bisogna dire che lo Z8000 aveva, almeno sulla carta, caratteristiche architetturali superiori all'8086: da un certo punto di vista era una specie di super insieme dello 8086, ma consentiva di evitare le complicazioni del modello di memoria di quest'ultimo. Anche il 68000 aveva caratteristiche architetturali più convincenti dell'8086: presentava anzitutto uno spazio di memoria lineare e si annunciava più potente. Tuttavia sia lo Z8000 sia il 68000, si fecero aspettare troppo a lungo dopo l'introduzione dell'8086. Il 68000 arrivò sul mercato con quasi due anni di ritardo, concedendo all'8086 un vantaggio che si sarebbe dimostrato incolmabile.

Ma fu l'introduzione dell'8088 a cambiare le cose. Al suo interno l'8088 era sostanzialmente identico all'8086, ma presentava un bus esterno a 8 bit. Anche Motorola e Zilog annunciarono le versioni a 8 bit (MC68008 e Z8008) dei loro micro a 16, ma quando l'IBM costruì il suo Personal Computer, diventato noto come IBM-PC, sul mercato era disponibile solo un micro a 8 bit capace di indirizzare almeno 1 MB di memoria: l'8088. Quello fu il punto di svolta dell'evoluzione del mondo dell'informatica. L'architettura Intel divenne lo standard di fatto nel mondo dei Personal Computer, la Microsoft stabilì le basi per diventare la più grande società di software del pianeta. Di questi sviluppi si parla con maggiori dettagli nell'Appendice G, dedicata all'architettura della famiglia ×86.

L'integrato MC68000.

Il 68000 ebbe ottima accoglienza da parte dei costruttori delle cosiddette *work-station*, ovvero sistemi individuali a carattere professionale, allora molto più avanzati dei calcolatori personali. Per essi il sistema operativo Unix era la norma. Il modello di memoria lineare del 68000 era adeguato per la tecnica di gestione

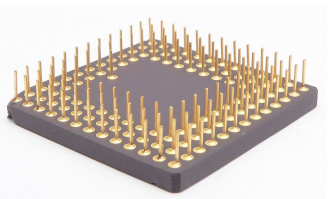
[13]La AMD strinse inizialmente un accordo con la Intel per la produzione dell'8086. Successivamente l'accordo non venne rinnovato. Le due società hanno continuato per anni a tenere rapporti ora litigiosi, ora collaborativi. La AMD ha dimensioni (e bilanci) molto inferiori a quelli di Intel, ma è saldamente il primo produttore di CPU compatibili. Ci sono stati periodi in cui le prestazioni delle CPU AMD hanno sopravanzato quelle delle corrispondenti CPU Intel.

della memoria virtuale di Unix. Il 68000 aveva inoltre più potenza dell'8086, anche perché bus dati e indirizzi non erano condivisi come per quest'ultimo (l'integrato MC68000 aveva 64 piedini, contro i 40 dell'8086). Aveva una architettura interna a 32 bit, indirizzi su 32 bit ma il bus dei dati era limitato a 16 bit. In altre parole era stata progettata una CPU a 16 bit nell'ottica di un prossimo passaggio a 32. Per anni l'architettura Motorola sarà lo standard di fatto delle più diffuse workstation. Il 68000 e i suoi successori sarebbero stati impiegati da Sun Microsystems, da HP, da Apollo (poi assorbita da HP) e da molti altri. Il 68000 venne pure impiegato nei Macintosh della Apple. La famiglia 68000 ebbe successo anche in campo industriale, nelle cosiddette applicazioni *embedded computer* richiedenti buone capacità elaborative, ad esempio nelle stampanti postscript.

Lo Z8000 non sfondò nel mercato dell'informatica dei grandi numeri, né trovò impiego in ambiente industriale. Venne impiegato nel Commodore, un calcolatore personale venduto essenzialmente come videogioco. In Italia, l'Olivetti usò lo Z8000 per il suo primo PC (denominato M20). L'M20 era una buona macchina, ma non incontrò i favori del grande pubblico – le attenzioni erano rivolte solo al PC IBM. Dopo poco tempo, Olivetti fu costretta a riconvertire la sua linea di calcolatori personali e passare alla produzione di macchine compatibili col PC: la torta era tanto ricca che un'esigua fettina valeva molto di più del totale dominio in aree parallele.

Con l'affermarsi dei microprocessori a 16 bit apparvero sul mercato svariati componenti di corredo avanzati, come gestori di memoria (MMU, *Memory Management Unit*), unità aritmetiche in virgola mobile (FPU, *Floating Point Unit*) e controllori di accesso alla memoria (DMA, *Direct Memory Access*). Il maggior problema che i costruttori si trovarono a fronteggiare non era più, a quel punto, la limitazione del numero di transistori, ma la potenza dissipabile dal chip. Per limitare tale dissipazione (che superava il watt), molte compagnie progettarono dispositivi in tecnologia CMOS. La tecnologia CMOS e l'abbandono delle precedenti tecnologie (TTL, ECL, ...) è stato il punto di partenza per gli impressionanti sviluppi futuri.

Nel 1984 Motorola introdusse l'MC68020, la prima effettiva CPU a 32 bit. Essa era realizzata in tecnologia CMOS da 1,5 μm e poteva arrivare a frequenze massime di funzionamento di 25 MHz. Nell'immagine accanto, il dispositivo viene mostrato dal lato dei piedini. Si noti l'elevato numero e la conseguente, più compatta organizzazione, detta *Pin Grid Array* (PGA). Il dispositivo aveva sia il bus dati sia il bus indirizzi a 32 bit e conteneva all'interno una piccola cache istruzioni. Era inoltre stato previsto un dispositivo esterno (MMU) per la gestione della memoria virtuale.



L'integrato MC68020.

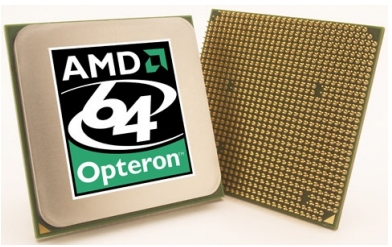
Intel arrivò dopo Motorola, nel 1985, a introdurre sul mercato la sua CPU a 32 bit in tecnologia CMOS. Si trattò dell'80386. Il 386 ha rappresentato una tappa fondamentale nell'evoluzione dell'architettura ×86. Non aveva cache interna, ma aveva integrata la gestione della memoria virtuale. Il modello di memoria virtuale dell'80386 si protrarrà, sostanzialmente immutato, fino ai nostri giorni. Successivamente (1987) Motorola presentò il 68030, simile al precedente ma con l'aggiunta di una memoria cache per dati e della MMU integrata. Per qualche tempo sembrò che il 68030 riuscisse a contrastare l'espansione dell'antagonista. Ma il mondo del PC era ormai saldamente in mano a Intel in virtù del patrimonio software esistente, mentre l'architettura 68000 aveva praticamente monopolizzato il settore delle workstation di buon livello.

Nel 1987 il 90% dei PC usava CPU Intel/AMD, mentre il 90% delle workstations impiegava CPU Motorola. Ma a lì a poco la situazione sarebbe cambiata, non tanto

per Intel/AMD, che sostanzialmente mantengono anche al giorno d'oggi la stessa quota nel mercato dei PC, quanto per Motorola che vide ridursi l'impiego dei suoi processori quando i maggiori produttori di workstation si volsero alle architetture RISC, che allora facevano la loro comparsa. In particolare, la Sun Microsystems, che aveva mosso i primi passi con l'architettura Intel ed era poi passata a quella Motorola, abbandonò anche quest'ultima per passare all'architettura RISC da essa sviluppata, denominata SPARC (*Scalable Processor Architecture*). Il motivo del passaggio era ovvio: eliminare la dipendenza dai fornitori esterni e ridurre i costi associati alle loro CPU, adottando processori i cui costi di produzione sarebbero stati inferiori.

Nel 1989 Motorola introdusse il 68040. Esso integrava 1.200.000 transistori in un singolo chip e, a parità di clock, aveva prestazioni 10 volte superiori al suo predecessore. L'anno successivo l'Intel presentò il suo nuovo microprocessore, l'80486, presto indicato da tutti più brevemente come 486. Nonostante l'uscita successiva al 68040, il 486 non fece altro che riconfermare il predominio dell'architettura ×86 nel settore dei calcolatori personali. Il 486 aveva prestazioni di riguardo, avendo a bordo una non trascurabile cache e l'unità in virgola mobile (non presenti sul suo predecessore 80386).

Di lì a qualche anno, la Motorola avrebbe praticamente cessato di sviluppare la linea 68000, dedicandosi allo sviluppo dell'architettura PowerPC, in compagnia di Apple e IBM. Nel seguito l'Intel ha continuato a sviluppare processori con architettura ×86 sempre più potenti. Nel 1993 introdusse il Pentium[14]. Successivamente, l'architettura ×86 venne estesa a 32 bit e ribattezzata dalla stessa Intel come IA32 (*Intel Architecture 32 bit*).



La CPU Opteron.

Ma è stata AMD, nel 2003, a immettere sul mercato il primo microprocessore (denominato Opteron, di cui un esempio di lato) che estendeva a 64 bit l'architettura ×86. In quell'arco di tempo Intel aveva scelto la strada di sviluppare una nuova e differente architettura a 64 bit, preannunciata come IA64. L'enorme successo dell'Opteron fece lievitare le vendite di AMD e fu gioco forza per Intel percorrere la stessa strada di AMD, cioè estendere a 64 bit l'architettura ×86. AMD adotta la denominazione AMD64, Intel adotta la denominazione Intel64[15].

I computer correnti, sia da tavolo sia portatili, contengono al loro interno quasi esclusivamente processori Intel o AMD. Data l'importanza dell'architettura ×86, questo libro la adotta come principale riferimento nei casi in cui ci si debba appoggiare a esempi di macchine concrete.

Il lettore avrà notato che il numero dei piedini dei microprocessori è andato aumentando. Il 4004 aveva 16 piedini, un microprocessore di alta gamma della produzione corrente (2016) ne ha più di un migliaio. Per consentire un tale numero di piedini si è passati dalla forma a spillo (*pin*) ai contatti sul retro dell'integrato, come appare nel-

[14]Con il Pentium Intel abbandonò la precedente convenzione di identificare le sue CPU con un numero. Ciò perché un tribunale non aveva accolto la sua richiesta di considerare i numeri come marchi di fabbrica e ciò consentiva ai suoi concorrenti, leggi AMD, di produrre processori contenenti gli stessi numeri di Intel. Il termine Pentium è una commistione di greco e latino, che voleva veicolare l'idea di quinta generazione, ovvero "architettura P5".

[15]La sigla IA64 viene riservata per denominare le CPU Itanium, sviluppate da Intel in collaborazione con HP agli inizi del 2000. L'Itanium presenta molte innovative soluzioni architetturali, tuttavia esso non ha ottenuto il successo che i suoi proponenti avevano sperato.

l'immagine dell'Opteron. I costruttori forniscono speciali zoccoli (*socket*), che vengono saldati sulla piastra e che presentano una corrispondente distribuzione di contatti. Lo zoccolo ha una parte meccanica che tiene fermo il processore consentendo il montaggio o la rimozione del medesimo. Questa è la soluzione adottata, per esempio, per consentire espandibilità dei sistemi a multiprocessore: si montano più zoccoli in modo da consentire l'aggiunta di processori negli zoccoli vuoti, quando ci sia la necessità di maggior capacità elaborativa. La produzione di nuovi modelli di processore si accompagna normalmente alla produzione dei corrispondenti zoccoli.
Un'altra soluzione è quella del montaggio superficiale (*surface mount*). In questo caso si parla di contatti in forma di *Ball Grid Array* (BGA). I contatti sono ricoperti da un saldante e, tramite un processo di riscaldamento, si determina il collegamento con i corrispondenti punti di contatto sulla piastra madre. Le ridotte dimensioni dei BGA permettono un maggior numero di contatti, anche se complicano il montaggio. La saldatura superficiale è usata tipicamente nei sistemi portatili, dove lo spazio a disposizione è limitato.

1.1.4 I microprocessori RISC

Le CPU di cui si è parlato in precedenza sono, o erano, macchine CISC (*Complex Instruction Set Computer*), ovvero macchine dotate di un repertorio di istruzioni molto ampio, come quasi tutte le macchine concepite negli anni settanta, quando il ricorso alla microprogrammazione era il metodo usuale nella progettazione delle unità di controllo.

Però c'era chi, già negli anni settanta, aveva cominciato a studiare differenti soluzioni architetturali, a seguito delle quali si è arrivati alle cosiddette macchine RISC (*Reduced Instruction Set Computer*). Più avanti nel libro, la discussione e il confronto tra le due differenti filosofie di progetto riceveranno il dovuto approfondimento; per il momento vengono forniti alcuni cenni storici su come queste ultime si sono sviluppate.

La prima macchina RISC venne realizzata in IBM, essenzialmente come prototipo di studio, e venne denominata IBM 801. Era costruita in logica discreta (logica ECL, molto veloce) e presentava 32 registri di 32 bit, aveva una cache istruzioni e una cache dati separate. Il prototipo, sviluppato presso il centro di ricerca di Yorktown Heights (NY), apparve nel 1979. Da questo prototipo IBM trarrà il modello 6150, immesso sul mercato nel 1986, che però avrà poco successo.

Gli studi relativi al modello 801 fecero da catalizzatore a un'ampia attività di ricerca: vennero sviluppati due prototipi da due Università della California, il RISC I, a Berkeley nel 1982, e il MIPS a Stanford, nel 1983; sostanzialmente due progetti nati nel contesto di corsi universitari.

Il RISC I ebbe una forte influenza sulla definizione della architettura SPARC da parte di Sun Microsystem. La prima CPU SPARC apparve nel 1987 e venne subito impiegata nelle workstation della Sun medesima, in sostituzione dei processori 68000. Il progetto originale era per una macchina a 32 bit, ma col tempo c'è stata la naturale evoluzione verso i 64 bit. Nel 2010 la Sun Microsystem è stata acquistata dalla Oracle Corporation. Oggi i processori SPARC trovano principalmente impiego nei *server* di elevate prestazioni ed alta affidabilità. L'immagine di fianco mostra una recentissima CPU SPARC prodotta da Oracle proprio per i server di alte prestazioni. È da rimarcare che nel 1988 è nata una associazione no-profit, di nome SPARC International, con l'obiettivo di promuovere lo svilup-

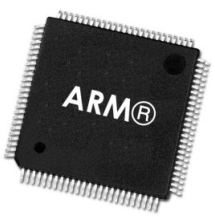
SPARC-M7

po dell'architettura e garantire la conformità alle specifiche dei processori immessi sul mercato dalle molteplici industrie licenziatarie[16]. La SPARC international ha reso liberamente disponibili le specifiche di alcuni modelli di CPU, proprio nell'intento di favorirne la diffusione.

A seguito del prototipo MIPS, nacque una società, che si chiamò MIPS Computers, che trasformò quel prototipo in un prodotto industriale indirizzato al mercato delle workstation grafiche e a quello dei sistemi *embedded*. Nel 2013 la società è stata acquistata da una industria inglese (Imagination Technologies), che oltre a produrre direttamente i processori, ne licenzia i progetti a industrie che intendono sviluppare specifici dispositivi attorno ai *core* MIPS. La pipeline del Capitolo 9 si ispira a quella definita originariamente per il MIPS.

Ma il RISC I influenzò anche la definizione di un'altra architettura di grande successo: l'ARM.

Oggi, la sigla ARM viene interpretata come *Advanced RISC Machine*, ma originariamente significava *Acorn RISC Machine*, dove Acorn era il nome di una società inglese costruttrice di personal computer, nata nel 1978, che nel 1983 decise di sviluppare una propria architettura. La ditta aveva costruito in precedenza un PC basato sulla CPU 6502[17]. Questo calcolatore venne chiamato BBC[18], come la nota rete televisiva inglese che in una serie di trasmissioni aveva indagato sul futuro dei "*microprocessor computers*" – allora al loro esordio – promuovendo lo sviluppo di un PC a basso costo per scopi didattici. Quando la Acorn decise di passare a una CPU piu potente del 6502, non trovò adatti per i suoi obiettivi i processori della classe 8086 e, essendo venuta a conoscenza della ricerca svolta a Berkeley, decise nell'imbarcarsi nella produzione di un suo microprocessore RISC. Ufficialmente il progetto venne lanciato nel 1983. La prima versione (oggi conosciuta come versione v1) ebbe luce nel 1985. Il dispositivo integrava 25.000 transistori e aveva prestazioni pari, se non superiori, a quelle dell'80286 apparso qualche anno prima, che però integrava molti più transistori (134.000) e presentava consumi molto più alti. Fu questa, in assoluto, la prima macchina RISC commerciale ad essere immessa sul mercato. ARM è stata sin dall'inizio a 32 bit; correntemente viene prodotta anche in versione 64 bit. Determinante per l'affermarsi dell'architettura ARM fu la scelta di Apple di utilizzarla nel suo PDA (Newton) nel 1990. Apple, Acorn e VLSI Technology svilupparono assieme la famiglia ARM6.



ARM-Cortex-M3

Essenzialmente ARM sta sul mercato vendendo tecnologia. Essa possiede la proprietà intellettuale (IP)[19] dell'architettura e cede la licenza d'uso agli effettivi produttori (tra questi Intel, Texas Instrument, Samsung, Apple e altre rinomatissime società). Questo modo di fare industria – senza disporre di un apparato industriale, fabbriche o simili – ha assunto notevole importanza negli anni recenti ed è in continua espansione, in particolare per la produzione dei cosiddetti SoC (*System-on-a-Chip*), specializzati per settori applicativi come quello dei telefoni mobili. La caratteristica

[16]Tra queste merita di essere ricordata la Fujitsu Microelectronics, che a suo tempo produsse la prima implementazione dello SPARC.

[17]Era questa la CPU utilizzata nei primi Apple.

[18]Il BBC, come il primo Apple, aveva la forma di una voluminosa tastiera al cui interno si trovava la logica di macchina e alla quale si poteva collegare un monitor (non grafico) e altri dispositivi, come le unità a cassette (analogiche).

[19]A volte si parla di SIP, ovvero *Silicon Intellectual Property*.

principale dell'architettura ARM è che il core del processore è estremamente semplice, confrontato con quelli dei processori di uso generale di grande diffusione, e contiene un numero di transistori molto basso. Ciò lo rende fabbricabile a basso costo e lascia libera sul chip un'area utilizzabile per implementare specifiche funzionalità.

Ogni costruttore è messo nella condizione di produrre un processore ARM arricchito delle funzionalità mirate a specifiche applicazioni (per esempio, aggiungendo un digital signal processor per le applicazioni che richiedono elaborazione del segnale). Dalla semplicità derivano anche bassi consumi. Tanto per dare un'idea, il processore PXA255 (facente parte della famiglia XScale sviluppata da Intel) a 400MHz fornisce prestazioni comparabili con quelle di un Pentium II a 300MHz, ma consumando 50 volte meno energia. Per questi motivi i processori ARM sono predominanti nelle applicazioni embedded e, in particolare, in quelle dove il basso consumo è un requisito primario, come per esempio nella telefonia mobile e nei tablet. Con ogni probabilità, il telefonino di chi sta leggendo queste righe ha al suo interno un processore ARM! Quanto a numeri di processori prodotti, l'architettura ARM non teme concorrenti: a Novembre 2015, sui siti ARM si leggeva che il numero complessivo di processori venduti aveva superato i 50 miliardi.

A conclusione della precedente carrellata, vale la pena ricordare che agli inizi degli anni novanta si formò un consorzio tra Apple, IBM e Motorola (denominato AIM) per la progettazione e la produzione di una nuova CPU di tipo RISC. Si trattava di un accordo che aveva un obiettivo dichiarato: battere sul campo dell'informatica di larga diffusione lo strapotere dell'architettura ×86. La nuova architettura si chiamò *PowerPC*. Essa si basava sulla precedente architettura di IBM denominata *Power*. Sulla carta, all'epoca in cui apparve, il PowerPC dimostrava di possedere quanto di meglio si potesse mettere assieme come idee e come realizzazione. Inizialmente le CPU PowerPC furono accolte molto favorevolmente dal mercato. Ovviamente sia Apple che IBM lo impiegarono nei loro sistemi. Nonostante ciò il PowerPC non è riuscito a contrastare il dominio dell'architettura ×86, sebbene sia stato impiegato per lungo tempo nei Macintosh di Apple. Dal 2006 Apple ha abbandonato i processori PowerPC a favore dei processori Intel. IBM ha continuato a produrlo e a usarlo in svariate linee di prodotto (compresi alcuni supercomputer).

Successivamente sono state attuate variazioni architetturali che hanno portato a quella che IBM definisce *Power Architecture* e, a partire dalla CPU Power4, un (*dual core*), l'IBM designa i propri processori semplicemente come Power (con di seguito un numero che ne individua il modello). Al momento di scrittura di questo testo il microprocessore di punta è il Power8. Queste macchine vengono normalmente impiegate per la costruzione di sistemi server di elevate prestazioni operanti sotto il sistema operativo AIX, versione IBM di UNIX. A partire dal 2013 IBM ha promosso la costituzione di un consorzio denominato *OpenPOWER Foundation*. L'obiettivo è quello di espandere il mercato di questi processori, rendendo aperte le specifiche architetturali e implementative, concedendo ai produttori la licenza di realizzare proprie versioni, secondo una modalità non dissimile da quella descritta a pagina 14 in riferimento ai core ARM.

La Tabella 1.1 riporta un elenco di microprocessori titolari di un qualche "primato". Non c'è alcuna pretesa di completezza, come, del resto, non ha pretesa di completezza la storia che abbiamo fin qui raccontato.

Caratteristiche principali di alcuni microprocessori			
Anno	Nome	Costruttore	Caratteristica
1971	4004	Intel	Prima CPU (1971)
1972	8008	Intel	Prima CPU a 8 bit
1974	8080	Intel	Prima CPU a 8 bit a grande diffusione
1974	6800	Motorola	Prima CPU con alimentazione singola (5V)
1974	PACE	National Inst.	Prima CPU a 16 bit *single chip*
1974	1802	RCA	Prima CPU in tecnologia CMOS
1976	8048	Intel	Primo microcalcolatore a 8 bit *single chip*
1978	8086	Intel	Prima CPU a 16 bit indirizzare oltre 64 KB
1979	8088	Intel	Prima CPU a 8 bit con architettura interna a 16 bit
1979	68000	Motorola	Prima CPU architettura (interna) a 32 bit
1984	68020	Motorola	Prima CPU con una *cache* istruzioni integrata
1985	Z80000	Zilog	Prima CPU con *cache* dati e istruzioni integrate
1985	ARM v1	ARM	Prima CPU RISC
1986	80386	Intel	Prima CPU a 32 bit con MMU integrata
1987	68030	Motorola	Prima CPU con integrate *cache* separate e MMU
1987	SPARC	Sun	Prima CPU RISC SPARC
1989	80486	Intel	Prima CPU con *cache*, MMU e FPU integrate
1992	601	AIM	Prima CPU PowerPc
1993	Pentium	Intel	Prima CPU con architettura superscalare
2001	Power4	IBM	Prima CPU *dual core*
2003	Opteron	AMD	Prima CPU (compatibile $\times$86) a 64 bit
2007	Core2 Quad	Intel	Prima CPU Intel *quad core*

Tabella 1.1 Caratteristiche distintive di alcuni microprocessori titolari di un qualche primato. La tabella tiene conto solo dei processori a grande diffusione e non considera alcune CPU che, pur essendo state introdotte per prime con qualche caratteristica di rilievo, non hanno avuto successo sul mercato.

1.2 Classi di calcolatori

Qui di seguito si esaminano le principali classi di appartenenza dei sistemi correnti.

1.2.1 Sistemi embedded

Il termine *embedded* è traducibile come *incorporato*. Per *sistema embedded* si intende un qualunque sistema il cui scopo di uso non è di per sé l'elaborazione dati, ma qualunque funzione applicativa. Nei sistemi embedded, un processore svolge la funzione di controllo, in sostituzione di apparati tradizionali (meccanici o altro), gestendo tutte o parte delle funzionalità dell'apparato in cui è immerso.

I sistemi embedded sono parte della nostra esperienza quotidiana e coprono un campo di applicazioni molto esteso: dal controllo di una lavastoviglie o di una bilancia elettronica al controllo di sistemi di guida di missili, di frenata con ABS, al controllo radar e via elencando. Corrispondentemente, il software può andare da qualche centinaio a qualche milione di righe di codice. Di norma il software è non modificabile e i programmi stanno in memorie di sola lettura, anche se i sistemi più complessi possono essere equipaggiati in modo tale da consentire l'aggiornamento del software stesso, con il caricamento delle

nuove versioni attraverso il sistema di comunicazione; in questi casi è naturale ricorrere alle memorie flash per memorizzare i programmi (oltre che i dati)[20].

In conseguenza dell'ampio spettro di impiego, vengono usati processori da 8, 16, 32 e 64 bit, i cui costi possono andare da pochi centesimi a qualche centinaio di Euro. Nei sistemi complessi, come quelli di guida, spesso il sistema embedded è costituito da una rete o da una gerarchia di processori comprendente processori di basso livello, direttamente collegati a sensori e attuatori, e processori di alta potenza elaborativa con funzioni di calcolo o di coordinamento dell'intero sistema. Un requisito frequente per i sistemi embedded è quello di operare in tempo reale. In alcuni casi ci sono vincoli di tempo reale stretto, che non possono essere infranti, a pena di conseguenze catastrofiche; si pensi ad esempio ai sistemi di guida automatici o ai vettori senza guida umana. In tal senso esiste un certo mercato per i sistemi operativi di tempo reale (RTOS, *Real Time Operating System*), ma non è infrequente lo sviluppo di software applicativo *stand alone*, che di sistemi operativi non fa uso.

Una sottoclasse dei sistemi embedded è quella degli apparati mobili personali, telefoni e *tablet*. Essi pure sono soggetti a vincoli di tempo reale, ad esempio nella presentazione delle immagini video sullo schermo di un tablet. Si tratta però di vincoli che, nelle usuali condizioni di uso, se non rispettati non hanno conseguenze catastrofiche, salvo l'insoddisfazione dell'utente. Questa categoria di sistemi funziona normalmente sotto il controllo di sistemi operativi di tutto rispetto (Android, iOS), che poco o niente hanno da invidiare ai sistemi operativi per calcolatori personali. Le applicazioni (App) che girano sui questi sistemi hanno raggiunto livelli di complessità e prestazioni confrontabili con quelli dei programmi dei sistemi più grandi. Alla data di scrittura di queste righe (2016), la maggior parte degli *smart phone* e dei tablet impiega processori a 32 bit, ma la tendenza verso processori a 64 bit sembra ormai essersi stabilita, almeno per i tablet.

Com'è ovvio, un fondamentale requisito per i processori impiegati nei dispositivi mobili è il basso consumo. Attualmente, l'architettura ARM la fa da padrone. Sostanzialmente essa copre oltre il 95% del mercato. Questi processori vengono prodotti in forma di SoC (*System on a Chip*) su licenza ARM da diverse industrie (Qualcomm, Apple, Samsung, Texas Instruments, Marvell, ..)[21]. Il grande successo di questa architettura è dipeso proprio dal basso consumo che ne ha determinato l'adozione da parte dei costruttori degli smart phone di maggior successo, dandole un vantaggio che oggi sembra incolmabile.

1.2.2 Sistemi portatili e da tavolo

Questa è la classe di sistemi che si è imposta con l'avvento dei microprocessori e che corrisponde al maggior segmento di mercato. Si possono acquistare sistemi in uno spettro di costo che va da sotto i 1.000 fino ai 3.000 Euro. Quanto a numeri di unità vendute c'è da dire che prima del 2008 i calcolatori da tavolo superavano i portatili, nel 2008 si è raggiunta una sostanziale parità, negli anni successivi la vendita dei portatili ha superato quella dei sistemi da tavolo. Nel frattempo i tablet hanno rosicchiato un buona fetta del mercato, e dal 2013 hanno superato il numero di unità vendute dei portatili.

[20] Vengono prodotti microcontrollori, ad esempio gli AVR di Atmel, che presentano integrate sullo stesso *chip* della CPU una memoria flash, come memoria di programma, oltre a memoria RAM e EEPROM.

[21] Si tratta di un mercato molto dinamico e difficile da seguire, perché, ad esempio, la Apple definisce il SoC, basato su ARM, per una propria linea di prodotti, ma la produzione fisica del relativo dispositivo è affidata all'esterno a una "fonderia di silicio", non esclusa la rivale Samsung.

1.2.3 Server

A pagina 7 abbiamo accennato al fatto che negli anni settanta del secolo scorso i mainframe e i minicalcolatori erano le categorie di macchine di maggior diffusione. Il dominio dei mainframe si è esteso ben oltre gli anni ottanta. Ma i microprocessori in un ventennio, raggiunsero capacità elaborative comparabili con quelle dei mainframe e in poco tempo rivoluzionarono il mercato. Tanto per avere dei punti di riferimento, si consideri che la CPU dell'IBM S/360-91, la macchina più potente della famiglia S/360 (fine anni sessanta) eseguiva 16 milioni di operazioni tra interi al secondo; il 486 (fine anni ottanta) era dato per 20 milioni di istruzioni a secondo. Il primo occupava armadi, il secondo pochi centimetri cubi.

Nello stesso arco di tempo c'è stata una profonda trasformazione nel mondo dell'informatica: la disponibilità di capacità elaborativa a basso costo ha favorito la realizzazione di sistemi basati sul concetto di *client-server*. In questo scenario l'uso della parola mainframe è andato via via scemando. Internet ha giocato un ruolo essenziale: la standardizzazione dei *browser* e dei protocolli di comunicazione ha reso comuni le cosiddette applicazioni web (*web application*), nelle quali i client sono per l'appunto normali browser web, mentre le applicazioni funzionano sui server (remoti) connessi via Internet. La stessa IBM classifica come *server* o, meglio, come *mainframe server* le macchine della "zSeries", punto di arrivo dell'evoluzione dell'architettura S/360.

I server sono normalmente costituiti da più di un processore, in modo da formare raggruppamenti (*cluster*) collegati in rete. Google, ad esempio, dispone di svariate installazioni (*Data Center*), molte in America, ma anche sparse per il mondo, dove sono dislocate centinaia di migliaia di server, macchine che tendenzialmente impiegano gli stessi processori dei computer da tavolo o portatili, sebbene sviluppati per essere usati in funzione di server (tanto per esemplificare, Intel chiama Xeon i processori destinati a impiego sui server).

I sistemi server devono garantire prestazioni adeguate in termini di transazioni elaborate nell'unità di tempo. Devono garantire un alto grado di scalabilità, ovvero possibilità di essere estesi in modo da fare fronte alla sempre crescente domanda di prestazioni; ciò normalmente si traduce nella facilità di aggiungere macchine a quelle esistenti quando il carico supera la capacità dei sistemi un uso. Infine, devono poter operare 24 ore su 24, con livelli di disponibilità elevatissimi; si parla del 99,99% (regola dei quattro 9). Si tenga presente che una disponibilità del 99,5% equivale a circa 44 ore di fermo all'anno. In Internet si trovano alcuni studi che mostrano le perdite cui una azienda può andare incontro a causa dei fermi.

I sistemi server trovano naturale impiego nel *cloud computing*.

1.2.4 Sistemi ad altissime prestazioni

I supercomputer sono macchine speciali, costruite per specifici scopi di ricerca e nell'intento, ovviamente, di servire in applicazioni ad altissima intensità di calcolo, ma anche per dare lustro alla nazione che li ospita e ai costruttori che li realizzano.

Un'organizzazione denominata Top500 dal 1993, per due volte all'anno, a Giugno e a Novembre, presenta la classifica dei 500 computer più potenti al mondo. La classifica è determinata in base al *benchmark* Linpack, che misura il numero di operazioni in virgola mobile eseguite al secondo. Il Linpack è essenzialmente un programma che risolve un sistema di equazioni lineari[22]

[22] Il benchmark Linpack è descritto in un *report*, intitolato *Performance of Various Computers Using*

La Tabella 1.2 riporta la classifica dei primi 9 computer a Novembre 2016. La Tabella 1.3 riporta quali sono stati i primi in classifica da Novembre 2009. Per ragioni di spazio le due tabelle sono necessariamente una sintesi di quelle originali.

Nome	Costruttore	Paese	N. totale di core	N. totale core cop	Freq GHz	Peta flops	Pot MW	Mflops/ Watt
TaihuLight	NRCPC	Cina	10.649.600	0	1,45	93,01	15,4	6051,3
Tianhe-2	NUDT	Cina	3.120.000	2.736.000	2,2	33,86	17,8	1901,5
Titan	Cray	USA	560.640	261.632	2,2	17,59	8,2	2142,8
Sequoia	IBM	USA	1.572.864	0	1,6	17,17	7,9	2176,6
Cori	Cray	USA	622.336	0	1,4	14,01	3,9	3557,9
Oakforest	Fujitsu	Giapp.	556.104	0	1,4	13,55	2,7	4985,7
K Computer	Fujitsu	Giapp.	705.024	0	2,0	10,50	12,7	830,2
Piz Daint	Cry	Svizz.	206.720	170.240	2,6	9,78	1,3	7453,5
Mira	IBM	USA	786.432	0	1,6	8,60	3,9	2176,5

Tabella 1.2 I primi 9 supercomputer del mondo a Novembre 2016, elencati in base alle prestazioni. I dati sono presi dal sito `www.top500.org`. Le prime 4 posizioni sono immutate dal Giugno 2016. La quarta colonna dà il numero totale di core (processori principali più coprocessori), mentre la quinta colonna dà il numero dei core dei coprocessori (quando questi sono usati). La settima colonna riporta le prestazioni, in termini di Peta flops (1 Peta $= 10^{15}$), misurate con il benchmark Linpack; l'ottava la potenza assorbita; l'ultima le prestazioni per Watt. Si noti come le frequenze (colonna 6) siano relativamente basse, per l'ovvio motivo di contenere i consumi che, in ogni caso, sono sempre superiori a qualche Mega Watt.

Da - a	Nome	Tipo di processore	N. core	Freq GHz	N. totale core proc	Peta flops
11/2009 – 6/2010	Jaguar (Cry)	Opteron	6	2,6	224.162	1,80
11/2010	Thiane-1A	Xeon	6	2,93	186.368	2,60
6/2011 – 11/2011	K Computer	SPARC	8	2,0	705.024	10,50
6/2012	Sequoia	PowerPC	16	1,6	1.572.864	17,17
11/2012	Titan	Opteron	16	2,2	299.008	17,59
6/2013 – 11/2015	Tianhe-2	Xeon	12	2,2	384.000	33,86
6/2016 – –	TaihuLight	Sunway	260	1,45	10.649.600	93,01

Tabella 1.3 Sistemi in prima posizione della classifica Top500 col trascorrere degli anni. La prima colonna riporta il periodo in cui il sistema di seconda colonna è stato primo (alla data della scrittura del TaihuLight non è dato sapere quanto durerà la sua permanenza in prima posizione). La terza colonna riporta il tipo di processore principale con, alla destra, il numero di core e la frequenza. La penultima colonna riporta il numero totale di core dei processori di terza colonna.

In riferimento alle Tabelle 1.2 e 1.3, vale la pena di osservare quanto segue.

- Ormai tutti i processori usati nei supercalcolatori sono a 64 bit. Il sistema operativo è per tutti un derivato di Linux.

Standard Linear Equations Software, di J. Dongarra, scaricabile da Internet (vedere il Paragrafo 1.6). Il programma di benchmark è pure scaricabile da rete.

Il secondo, il sesto e il settimo sistema di Tabella 1.2 sono degli MPP (*Massively Parallel Processor*), ovvero dei sistemi costituiti da più processori strettamente interconnessi tra loro. I restanti sistemi di Tabella 1.2 sono dei *cluster*, ovvero raggruppamenti di processori parzialmente indipendenti, che utilizzano (per lo meno in forma di sottogruppi) distinte risorse (memoria, dischi, ecc.).

- A Novembre 2016, la macchina più potente al mondo è il Sunway TaihuLight, dislocata presso National Supercomputing Center in Wuxi, Cina. Questa macchina contiene ben 10.649.600 core, che derivano da 40.960 processori Sunway SW26010, interamente progettati e fabbricati in Cina dallo Shanghai High Performance IC Design Center. Ogni processore SW26010 ha al suo interno 4 gruppi di core. Ogni gruppo contiene 64 processori (detti *Computing Processing Element*, CPE) organizzati come una matrice 8×8 e un processore (detto *Management Processing Element*, MPE) che li coordina. Il totale è di 260 core per ciascun integrato SW26010. Tutti i singoli core sono processori RISC a 64 bit, in grado di effettuare operazioni vettoriali a 256 bit.
  Le prestazioni sono pari a 93,01 Petaflops ($93,01 \times 10^{15}$ operazioni in virgola mobile al secondo). Il consumo è pari a 15,4 MW, corrispondenti a 6051,3 Mflops/Watt.
- La colonna 4 dà il numero totale di core; la colonna 5 dice quanti di essi sono core di coprocessori aritmetici[23]. Si noterà che 3 dei 9 sistemi elencati operano con coprocessori distinti rispetto ai processori attorno a cui sono costruite le macchine. Sostanzialmente ci sono due linee: usare processori che dispongono di alte capacità elaborative (per esempio operazioni vettoriali) e non affiancare nessun coprocessore, oppure usare più convenzionali processori affiancando loro i coprocessori. Il Sunway TaihuLight segue la prima strada; il processore SW26010 dispone di tutta la capacità elaborativa che serve per il calcolo scientifico. Il Tianhe-2 e il Titan seguono la seconda strada; essi adoperano normali processori (ad uso server), corredati da un elevato numero di coprocessori.
  In particolare il Tianhe-2, (detto anche MilkyWay-2), dislocato presso il Centro per il supercalcolo di Guangzhou, Cina, è un cluster di processori *Core* Intel della terza generazione, denominata "Ivy Bridge"[24]. Più precisamente, Thiane-2 impiega 32.000 processori Xeon a 2.2 GHz, ciascuno dei quali contiene al suo interno 12 core, per un totale di 384.000 core di CPU, e 48.000 coprocessori Xeon Phi, ciascuno dei quali ha al suo interno 57 core, per un totale di 2.736.000 core di coprocessore; ciò porta a un totale complessivo di 3.120.000 core.
- Fino a pochi anni addietro le prime posizioni della classifica erano saldamente tenute da sistemi dislocati negli USA, salvo qualche eccezione di sistemi giapponesi e (pochi) sistemi europei. A partire da Giugno 2013 la testa della classifica è sempre stata cinese.
- Quanto a marche di processori usati, la Tabella 1.2 sembra indicare una situazione quasi equilibrata; in realtà l'Intel fa la parte del leone: delle 500 macchine elencate a Novembre 2016, 462 (92,0%) usano processori Intel, 22 (4,4%) usano processori Power o PowerPC IBM, 7 (1,4%) usano Opteron AMD, 7 (1,4%) usano processori SPARC. Negli anni precedenti il 2010 la situazione era molto più equilibrata e con un maggior numero di attori. Oggi, la situazione sembra volgere a netto vantaggio di Intel, salvo il

[23] Questi coprocessori sono GPU; di essi di parla al Paragrafo 11.2.3.

[24] Dal 2006 Intel designa i propri processori come "Core" e li classifica con un "nome in codice" (*Codename*) e la corrispondente generazione (entro la famiglia Core). Il nome in codice Ivy Bridge identifica la terza generazione dei processori Core. Il nome di codice Broadwell identifica la quinta generazione, introdotta a Gennaio 2015. Nella seconda metà del 2015 è stata introdotta la sesta generazione, denominata Skylake. Gli specifici modelli entro una generazione vengono denominati i3, i5 o i7, per le CPU di uso generale (alle tre sigle corrispondono prestazioni crescenti), oppure Xeon, per le CPU pensate per le macchine della categoria server.

Figura 1.2 A sinistra una vista del Sunway TaihuLight, a destra una vista delThiane-2.

primo della classe entrato di recente in classifica.

1.3 Uno sguardo all'evoluzione tecnologica

Nell'Aprile del 1965 Gordon Moore, che assieme a Robert Noyce e Andy Groove nel 1968 fonderà l'Intel, in un articolo sulla rivista *Electronics*[25] ipotizzò che l'incremento della capacità elaborativa sarebbe continuato per tutti gli anni 70 al ritmo di un raddoppio ogni 12 mesi. Moore aveva appena finito di realizzare un chip contenente 60 transistori, il doppio di quello che aveva realizzato l'anno precedente. Per "aumento della capacità elaborativa" egli, in realtà, intendeva "aumento del numero di transistori" nel singolo chip. Vedremo che le due cose non sono perfettamente equivalenti.

Questa previsione fu poi corretta dallo stesso Moore due volte: nel 1975 portandola a un raddoppio ogni 2 anni (ed estendendone la validità agli anni 80) e a fine anni 80 portandola a 18 mesi. In effetti dagli anni 80 al primo quinquennio del 2000 la velocità dei processori è cresciuta in accordo a ques'ultima previsione, quadruplicando ogni 3 anni (si veda più avanti la Tabella 1.5). La previsione di Moore è diventata il metro di misura e l'obiettivo per le aziende che operano nel settore, tanto da essere percepita come una legge, la cosiddetta *legge di Moore*.

Come è facile immaginare, più di uno scettico ha levato la sua voce contro la sostenibilità dell'evoluzione quantitativa prevista da questa "legge". Non c'è alcuna ragione per cui essa debba continuare a valere: non è una legge di natura, bensì una legge sull'ingegno umano. Prima o poi la previsione di Moore verrà smentita, ma gli scettici avranno dovuto aspettare un bel po' prima di vedere confermate le loro critiche[26].

La legge di Moore spiega in termini quantitativi, meglio di qualunque argomentazione a parole, lo spettacolare sviluppo dell'elettronica a cui si sta assistendo da anni. Affermare che la microelettronica raddoppia la sua capacità ogni 18 mesi significa affermare che nei

[25] All'epoca era una rivista di larghissima diffusione. Veniva distribuita gratis, anche in Italia, a chiunque fosse ritenuto un soggetto capace di influenzare acquisti di componenti o apparati elettronici.

[26] Per capire i limiti fisici teorici della legge di Moore, cioè quali sono la massima potenza elaborativa e la massima densità di informazioni raggiungibili, si può leggere una serie di articoli sul numero del 31 agosto 2000, vol. 406, della rivista *Nature* (rintracciabile su Internet), in particolare l'articolo "*Ultimate physical limits to computation*" di Seth Lloyd (pp.1047–1054). Più recentemente, a maggio 2015, su *Scientific American* è apparso un articolo, pure rintracciabile in Internet, dal titolo eloquente: "*Moore's Law Keeps Going, Defying Expectations*".

prossimi 18 mesi avremo un incremento equivalente a quello che si è avuto fino a oggi. Un tale tasso di evoluzione comporta delle conseguenze che verrebbe voglia di chiamare fantascientifiche, se esse non fossero parte della vita di tutti i giorni.

Anzitutto c'è una accentuata riduzione dei costi nel tempo, in netta controtendenza rispetto a qualunque produzione industriale. Si capisce perché un personal computer corrente ha una capacità elaborativa qualche centinaio di volte superiore a quella di un mainframe di quaranta anni addietro, nonostante che il primo costi intorno al migliaio di euro, mentre un miliardo di lire (di allora) poteva anche non bastare per un mainframe. Chi è interessato a volgere il suo sguardo al futuro troverà abbondante materiale in Internet. A noi interessa vedere più da vicino qual è il tasso di crescita dei principali componenti dei calcolatori.

1.3.1 Processori

Il diagramma di Figura 1.3, con riferimento ai processori Intel, mostra come è andato aumentando negli anni il numero dei transistori contenuti in successivi modelli di CPU a partire dal 4004 fino all'i7 nella versione di quinta generazione dei processori Core (terminologia Intel, vedere la Nota 24 a pagina 20). Si osservi che la scala dell'asse verticale è logaritmica. Il numero dei transistori viene pure riportato nella terza colonna di Tabella 1.4. La due colonne a destra della tabella rappresentano il livello di tecnologia e la frequenza di clock delle relative CPU.
Dai dati di Tabella 1.4 si scopre che in realtà, almeno per i microprocessori Intel, la crescita del numero di transistori è stata al disotto del raddoppio ogni 18 mesi[27].

I dati riportati nella Tabella 1.4 si riferiscono alle versioni di introduzione dei differenti modelli di CPU. Di norma, la versione di introduzione viene migliorata al rilascio di versioni successive. Per esempio, il Pentium 4, che nel 2000 integrava 42 milioni di transistori e aveva una cache di secondo livello sul chip di 256 KB, dal 2002 è stato dotato di una cache di 512 KB passando a 55 milioni di transistori; nel Febbraio 2004, venne prodotta una versione con una cache di secondo livello da 512 KB e una di terzo livello da 2 MB integrate sul chip, portando il conto complessivo dei transistori a 178 milioni.

Secondo la formulazione classica della legge di Moore le prestazioni sono direttamente legate al numero di transistori. Ma non è detto che ci sia una effettiva proporzionalità. Si faccia, per esempio, riferimento alla memoria *cache*. Una cache più ampia fa crescere le prestazioni, ma l'incremento di prestazioni che consegue da una cache più ampia di norma è meno che proporzionale all'aumento del numero di transistori richiesti per estenderla. Al Paragrafo 2.9.2 impareremo a dare una valutazione quantitativa dell'impatto del miglioramento di una parte del sistema sulle prestazioni complessive.

La Tabella 1.4 mostra che, assieme al numero di transistori contenuti nei chip delle CPU, sono andati aumentando anche la frequenza e la densità. L'evoluzione della microelettronica passa sia attraverso la progressiva riduzione delle dimensioni dei transistori sul chip, sia attraverso la riduzione delle dimensioni delle connessioni tra di essi. La velocità di elaborazione dipende, in primo luogo, da quanto velocemente è possibile azionare il singolo transistore e dal tempo impiegato a trasmettere l'informazione da un transistore all'altro. Minori dimensioni comportano minor tempi di latenza nella trasmissione dei segnali e, conseguentemente, la possibilità di accrescere la frequenza. La dimensione dei circuiti di connessione è utilizzata come metro di misura del livello di tecnologia raggiunto.

[27] La crescita è di circa 2,6 volte ogni 3 anni, corrispondente a un tasso di crescita del 40% annuo.

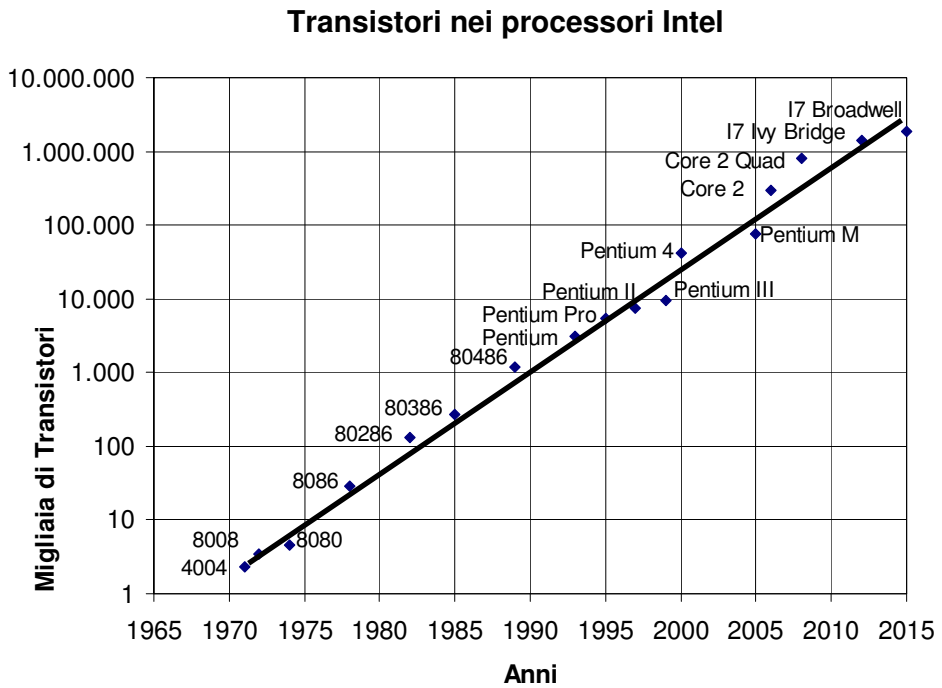


Figura 1.3 Aumento del numero dei transistori nelle CPU Intel.

Alla data di scrittura di queste pagine (2016), la produzione di avanguardia di Intel è a 0,014 μm.

Come vedremo al Paragrafo 2.9 la frequenza del *clock* è uno dei principali parametri che determinano le prestazioni. Tanto per fare un esempio, tra il Pentium 4 e l'8086 c'è un rapporto tra le frequenze pari a $1400/5 = 280$ (Tabella 1.4); ovvero, il primo ha prestazioni 280 volte superiori al secondo solo come conseguenza dell'aumento della frequenza, senza considerare i miglioramenti architetturali.

Dalla metà degli anni ottanta, fino a circa la metà del primo decennio del 2000, il tasso di crescita delle prestazioni è stato tra il 50% e il 60%, corrispondente a circa poco meno di una quadruplicazione ogni tre anni. Questo è stato il tasso di crescita dell'epoca in cui le prestazioni erano in larga parte ottenute aumentando la frequenza. La Tabella 1.4 mette in evidenza come la frequenza sia andata crescendo in modo cospicuo almeno fino a circa metà del primo decennio del 2000, ma poi si sia sostanzialmente stabilizzata. L'impossibilità ad accrescere ulteriormente la frequenza (di rado supera 3,7 GHz) non consente più di mantenere un tale tasso di crescita. A partire dal 2000 l'incremento delle prestazioni si è ridotto passando al 40% nel periodo 2000-2004 e passando al 20% negli anni successivi.

A partire dal 2006, Intel e i suoi concorrenti hanno iniziato a produrre i cosiddetti *multicore*, ovvero chip che integrano più nuclei (*core*) di elaborazione, contornati da risorse a comune. Le sigle "Duo" e "Quad", stanno a indicare la presenza di 2 o 4 core[28].

[28] Si veda anche l'Appendice C, dove si fa un resoconto dell'evoluzione dell'architettura ×86.

Data di introduzione	Nome del chip	N. transistori (/1000)	Tecnologia (μm)	Frequenza (MHz)
Novembre 1971	4004	2,3	10	0,108
Aprile 1972	8008	3,5	10	0,500
Aprile 1974	8080	4,5	6	2
Giugno 1978	8086	29	3	5
Febbraio 1982	80286	134	1,5	8
Ottobre 1985	80386	275	1,5	16
Aprile 1989	80486	1.200	1	25
Marzo 1993	Pentium	3.100	0,8	60
Novembre 1995	PentiumPro	5.500	0,6	150
Maggio 1997	Pentium II	7.500	0,35	233
Febbraio 1999	Pentium III	9.500	0,25	450
Novembre 2000	Pentium 4	42.000	0,18	1400
Marzo 2003	Pentium M	77.000	0,13	1300
Settembre 2006	Core 2 Duo E4300	291.000	0,065	1800
Ottobre 2008	Core 2 Quad Q8200	820.000	0,045	2330
Aprile 2012	i7-3770 3a gen. (Ivy Bridge)	1.400.000	0,022	3400
Gennaio 2015	i7-5557U 5a gen. (Broadwell)	1.900.000	0,014	3100
Settembre 2015	i7-6700HQ 6a gen. (Skylake)	1.750.000	0,014	2600

Tabella 1.4 Aumento del numero di transistori delle CPU Intel. I dati riportati si riferiscono al modello di introduzione.

Il passaggio alle versioni multicore è stato indotto dall'impossibilità di aumentare le frequenze oltre un certo limite a causa di problemi di dissipazione del calore. I multicore sono una soluzione architetturale per ottenere maggiori prestazioni; tuttavia la crescita di potenza è meno che lineare rispetto alla crescita del numero di core, a causa dei conflitti che si generano tra i diversi core nell'uso delle risorse comuni e a causa dei problemi relativi al coordinamento delle attività.

1.3.2 Memorie

In Figura 1.4 viene riportato l'incremento della capacità degli integrati di memoria DRAM, il tipo di memoria impiegato come memoria principale del calcolatore. Nell'arco di tempo dal 1980 a circa la fine del XX secolo, la capacità è andata crescendo di circa un 60% all'anno corrispondente a una quadruplicazione ogni 3 anni. Più recentemente si osserva un raddoppio ogni 2 anni
Per quanto riguarda le prestazioni, identificabili con il tempo di ciclo, esse hanno avuto una crescita molto più lenta, con un tasso di poco più del 7% annuo, corrispondente a un raddoppio ogni 10 anni. Nell'ultimo ventennio le DRAM sono diventate sincrone (SDRAM) si sfruttano ambedue i fronti del clock, con le cosiddette DDR SDRAM (*Double Data Rate Synchronous Dynamic Random Access Memory*). Le memorie DDR seguono gli standard definiti dal JEDEC, un consorzio cui partecipano i costruttori. A fine 2014 sono apparsi sul mercato i primi dispositivi a standard DDR4, originariamente attesi per il 2012, sebbene la loro specifica fosse iniziata più di un decennio prima. A fine 2015 sono

Anno	Tipo	Capacità (Mb)	Tempo di ciclo (ns)
1980	DRAM	0,064	250
1983	DRAM	0,256	220
1986	DRAM	1	190
1989	DRAM	4	165
1992	DRAM	16	145
1996	SDRAM	64	120
2000	DDR1	256	100
2002	DDR1	512	80
2004	DDR2	1000	70
2006	DDR2	2000	60
2010	DDR3	4000	40
2014	DDR4	8000	20

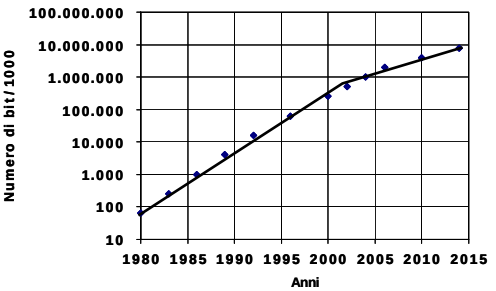


Figura 1.4 Aumento della capacità per gli integrati DRAM. La capacità è misurata in Mb, il tempo di ciclo in ns. Le seconda colonna indica il tipo, ovvero la generazione; DDR sta per *Double Data Rate (Synchronous Dynamic Random Access Memory)*.

annunciati integrati a 32 Gb e si sta parlando di "generazione post-DDR4", anche se per essa si prevede un ingresso nel mercato verso il 2020.

In Tabella 1.5 viene riportato, in estrema sintesi, il tasso di crescita osservato a partire dagli anni ottanta fino ai primi anni del 2000, relativamente ai tre componenti fondamentali di un calcolatore: CPU, memoria DRAM e memoria a dischi. La tabella serve a dare un'idea degli ordini di grandezza in gioco. Nell'arco di tempo in questione il tasso di crescita della velocità delle CPU è stato tra il 50 e il 60 percento, mentre quello delle memorie del 7% circa, con un differenziale di crescita superiore al 50% all'anno. Questa progressiva divaricazione è stata la ragione che ha motivato l'introduzione delle memorie *cache*. I dati contenuti in tabella sono da intendersi come indicativi, in particolare essi non riflettono esattamente la tendenza dell'ultimo decennio. A partire da circa il 2003 la velocità della CPU è andata crescendo di poco sopra al 25%; con tendenza a dimunuire col passare degli anni. Facendo riferimento alla Tabella 1.4 e alla tabella di Figura 1.4, si osserva che nell'ultimo decennio il numero di transistori non è andato raddoppiando ogni 18 mesi, mentre il tempo di accesso alla memoria DRAM ha subito una forte riduzione.

	Capacità		Velocità	
	Crescita	Tasso annuo	Crescita	Tasso annuo
Logica	2× in 3 anni	26%	4× in 3 anni	60%
DRAM	4× in 3 anni	60%	2× in 10 anni	7%
Dischi magnetici	4× in 3 anni	60%	2× in 10 anni	7%

Tabella 1.5 Tasso di sviluppo delle tre principali tecnologie impiegate nei sistemi di elaborazione nell'arco di tempo 1980-2000. I dati di questa tabella hanno puro valore indicativo. Peraltro, dalla metà del primo decennio degli anni 2000, il tasso di crescita della velocità delle CPU si è ridotto attorno al 25% all'anno, mentre, nello stesso decennio, la velocità delle DRAM è andata aumentando molto più del 7% annuo, come si osserva nella tabella di Figura 1.4.

1.3.3 La questione dei consumi

La Tabella 1.2 offre lo spunto per una riflessione sui consumi di energia elettrica. La settima colonna mostra che le potenze in gioco sono ragguardevoli. Sebbene questa tabella si riferisca a una categoria di macchine speciali, è facile intuire che il consumo di energia elettrica associato all'operatività dei sistemi di elaborazione non è un consumo marginale nei paesi tecnologicamente progrediti. Una delle problematiche del cosiddetto *Green Computing* è proprio quella di ridurre al minimo gli impatti ambientali che da tali consumi derivano.

C'è poi da considerare l'aspetto prettamente tecnologico, ovvero quello legato alla dissipazione del calore. Mentre i primi microprocessori consumavano meno di 1 Watt, i processori moderni possono arrivare anche a superare i 100 Watt. Con il ridursi delle dimensioni la dispersione del calore diventa problematica. In Figura 1.5 viene mostrata la densità di potenza in funzione della tecnologia. Come si vede già con il Pentium Pro si era raggiunto un livello comparabile con quello delle piastre dei fornelli elettrici e si stava avvicinando quello dei reattori nucleari (a fine 2015 la tecnologia Intel è a 0,014μm, Tabella 1.4).

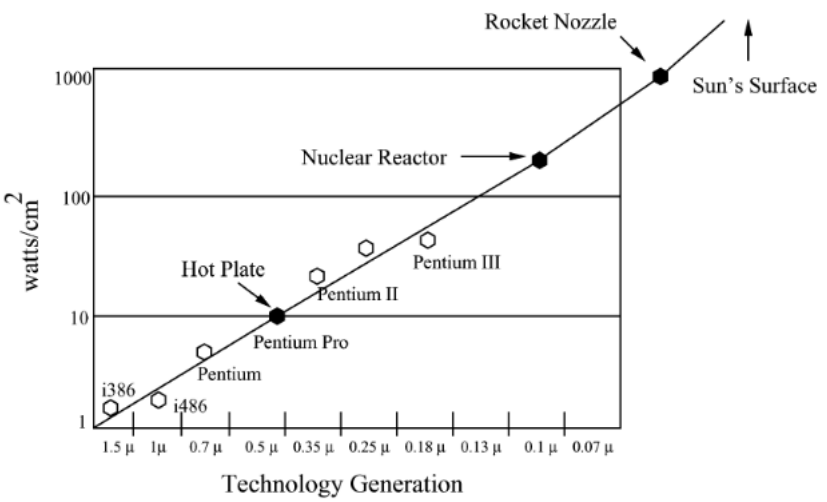


Figura 1.5 Aumento della densità di potenza in funzione della tecnologia. La figura è riportata direttamente da [VF05].

La potenza dissipata è in larga parte dovuta alla commutazione dei transistori (potenza dinamica) ed è data da

$$P_{dinamica} \sim CV^2 f \tag{1.1}$$

Dove C è la capacità, V la tensione e f la frequenza, [VF05].

La 1.1 suggerisce che il primo fattore su cui intervenire per ridurre la potenza dissipata è la tensione, dato che dipende da quest'ultima in modo quadratico. I primi microprocessori erano a 5V, oggi le tensioni sono attorno a 1V. La capacità C dipende dalla tecnologia (dimensione dei transistori e dei collegamenti); C si riduce con la riduzione delle dimensioni. La frequenza è il terzo parametro su cui giocare. Infine, la dissipazione

può essere ridotta togliendo il clock (*clock gating*) alle parti del processore che sono inattive. I tre parametri che compaiono nella 1.1 non sono del tutto indipendenti, nel senso che se per esempio si riduce la tensione di alimentazione si ha l'effetto di aumentare i ritardi e questo implica la necessità di ridurre la frequenza. Scalare assieme la tensione di alimentazione e la frequenza prende il nome di DVS (*Dynamic Voltage Scaling*). DVS è la tecnica più adottata essendo la riduzione ottenibile teoricamente cubica.

La Figura 1.5 e la 1.1 spiegano perché a partire dai primi anni 2000 le frequenze non sono sostanzialmente più aumentate. A causa della dissipazione del calore, i produttori hanno dovuto rinunciare agli aumenti di frequenza, rifugiandosi nello sviluppo di apparati multicore.

1.4 Architettura e organizzazione

Il concetto di *architettura* venne introdotto con il sistema IBM S/360 nel 1964. Riportiamo qui di seguito la definizione da [ABB64].

> *The term architecture is used here to describe the attributes of a system as seen by the programmer, i.e., the conceptual structure and functional behavior, as distinct from the organization of the data flow and controls, the logical design, and the physical implementation.*

In altre parole, con il termine *architettura di un calcolatore* si intende ciò che l'utente (il programmatore in linguaggio assemblativo) deve conoscere della macchina, in sintesi:

- il repertorio delle istruzioni;
- il formato dei dati;
- le modalità di indirizzamento e la disponiilità di registri.

Con il termine *organizzazione*, ci si riferisce alle relazioni strutturali tra le unità funzionali che compongono il calcolatore e al modo in cui esse realizzano una data architettura. L'organizzazione di un calcolatore è tendenzialmente non visibile al programmatore. Di essa fanno parte aspetti quali la tecnologia impiegata, l'eventuale presenza di una pipeline e il modo in cui viene eseguita una data istruzione.

Decidere se l'istruzione di moltiplicazione farà parte del repertorio di istruzioni è una questione architetturale, mentre decidere se per l'esecuzione di questa istruzione è prevista una specifica unità di moltiplicazione o se questa è ottenuta attraverso l'impiego dell'unità di somma comunque presente nella macchina è una decisione organizzativa. Decidere il numero di registri di uso generale di cui sarà provvista la macchina fa parte della progettazione architetturale, mentre decidere se l'unità di controllo sarà in logica cablata o microprogrammata fa parte dell'organizzazione.

Spesso, anziché di organizzazione si parla di *microarchitettura*. Per esempio, con riferimento ai processori i7 dell'architettura ×86, Intel designa le differenti generazioni con i nomi che ne identificano la microarchitettura, la prima (2006) si chiamò *Core*[29], la seconda (2008) Nehalem, l'attuale (2015) è denominata Skylake. Tutte implementano la stessa architettura i7.

Il lettore è comunque avvisato che la differenza tra i termini "architettura" e "organizzazione" è piuttosto sfumata. Per esempio, quasi tutti i processori correnti dispongono

[29] Il termine Core è rimasto per designare la famiglia in generale.

di una cache e della memoria virtuale e dispongono di istruzioni che servono a controllarne il funzionamento. Se in genere la cache e la memoria virtuale sono trasparenti al programmatore, e quindi non sarebbero da considerare come aspetti architetturali, per il programmatore di sistema la cache e la memoria virtuale entrano a buon diritto nell'architettura. Infine ci sarebbe da considerare l'aspetto dell'*implementazione* ovvero come una data architettura/organizzazione viene realizzata da un punto di vista dell'hardware. In questo libro si trattano sia gli aspetti architetturali sia quelli organizzativi sia quelli legati all'hardware, anche se preferenzialmente useremo il solo termine architettura.

A pagina 6 abbiamo detto che il concetto di architettura venne introdotto con il sistema IBM S/360. Per la prima volta il cliente poteva passare da un modello meno potente (e quindi meno costoso) a un modello più potente mantenendo (quasi) inalterato il software.

Il concetto di architettura avvantaggia il cliente salvaguardandone l'investimento per il software, avvantaggia il costruttore in quanto gli consente di adottare nuove tecnologie e nuove soluzioni organizzative fornendo prodotti più appetibili. La compatibilità architetturale è stata una delle chiavi del successo dei sistemi Intel. Con il microprocessore 8086, introdotto nel 1978, è stata definita l'architettura ×86 che, con aggiunte e miglioramenti, si è trasformata in quella chiamata IA32 e, più recentemente, Intel 64. Lungo questo percorso sono state impiegate nuove tecnologie microelettroniche e nuove strutture organizzative. Ma è stata mantenuta la compatibilità, nel senso che programmi eseguibili su un 8086 sono tendenzialmente eseguibili sui multi core correnti.

Modello di programmazione/*Instruction set architecture*
Nel testo viene usato il termine *modello di programmazione*. Con esso ci si riferisce a ciò che vede il programmatore assembler, ovvero a ciò che all'inizio di questo paragrafo è stato definito come architettura.
Alternativamente in letteratura è molto diffuso il termine ISA (*Instruction Set Architecture*). I due termini sono sinonimi.
In questo libro li useremo entrambi, anche se preferenzialmente faremo ricorso al concetto di "modello di programmazione".

1.5 Livelli e astrazioni

Nelle scienze esatte la realtà viene rappresentata attraverso modelli. Un modello costituisce sempre un'astrazione della realtà che intende rappresentare. Con esso si evidenziano gli aspetti di interesse e si sopprimono i dettagli non necessari.

Un moderno calcolatore costituisce un sistema estremamente complesso. Per poterne comprendere il funzionamento occorre far ricorso a metodi di descrizione appropriati, costruendo modelli che siano funzionali alla rappresentazione degli aspetti che si intende analizzare. Nello studio delle architetture dei calcolatori, come del resto in larga parte dei sistemi informatici, risulta conveniente una rappresentazione secondo livelli (di astrazione) come in Tabella 1.6.

La suddivisione in livelli concettuali consente di affrontare l'analisi e il progetto in modo ordinato. Ogni livello è costituito da un insieme di componenti e da un insieme di modi per combinare i componenti in strutture. Le strutture così costruite rappresentano l'astrazione realizzata dal livello e costituiscono le entità su cui si opera a livello superiore. Ogni livello nasconde i dettagli non necessari e costruisce l'insieme di astrazioni rilevanti per il livello superiore. Questo, a sua volta, le aggrega in modo da formare entità ancora

più astratte per il livello soprastante. Ogni livello è caratterizzato da linguaggi descrittivi e metodi analitici caratteristici del livello stesso.

Le astrazioni coincidono spesso, ma non necessariamente, con gli effettivi confini fisici del sistema, proprio perché esse vengono utilizzate dal progettista per poterne gestire la complessità.

Livello delle applicazioni	Strutture: programmi applicativi Componenti: sistema operativo, librerie, file system
Livello del linguaggio di macchina	Strutture: programmi (assembler) Componenti: modello di programmazione, repertorio istruzioni
Livello funzionale (RTL)	Strutture: unità di controllo, modello di programmazione Componenti: registri, bus, memorie
Livello della logica	Strutture: registri, contatori, unità aritmetiche memorie Componenti: porte, flip-flop, clock
Livello dei circuiti	Strutture: porte logiche, flip-flop, driver Componenti: transistori, resistenze, capacità

Tabella 1.6 Schematizzazione a livelli di un sistema di elaborazione. I livelli che interessano l'architettura e l'organizzazione dei calcolatori sono evidenziati in grigio.

.

Nello schema di Tabella 1.6, il livello più basso è quello che corrisponde ai circuiti. Qui si trovano componenti come transistori, resistenze, capacità ecc. Il livello struttura questi componenti in modo da formare entità come porte logiche, driver, flip-flop, ecc.

Il livello soprastante è quello della logica. I componenti di questo livello sono le porte, gli elementi di memoria (flip-flop), gli elementi di ritardo, ecc. Cioè le astrazioni realizzate dal livello dei circuiti. I componenti vengono strutturati in entità quali registri, contatori, unità aritmetiche, memorie, selettori (*multiplexer*), codificatori, decodificatori.

Il successivo livello è quello relativo ai trasferimenti tra registri. Il livello è perciò denominato RTL (*Register-Transfer Level*)[30]. Qui gli elementi componenti sono le entità appena menzionate e la loro strutturazione corrisponde alla logica di controllo. Il livello realizza l'astrazione corrispondente al modello di programmazione (formato dei dati, repertorio di istruzioni, indirizzamento ecc.).

Il livello seguente è quello del linguaggio di programmazione assembler. La macchina viene vista attraverso il suo modello di programmazione. La struttura è data dalla aggregazione di istruzioni e dati, in modo da formare dei programmi.

Il livello più alto è quello che corrisponde ai programmi applicativi. Questo livello si appoggia sulle astrazioni ottenute attraverso il livello precedente e cioè il sistema operativo, il file system e tutti i servizi resi disponibili per i programmi di utente.

La schematizzazione di Tabella 1.6 è alquanto arbitraria ed è essenzialmente mirata a mettere in mostra, in modo sintetico, gli aspetti di interesse di questo libro. La schematizzazione è anche alquanto grossolana. Infatti, una separazione netta tra confini come quella di figura non sempre è sufficiente: si pensi ad esempio alle applicazioni dell'utente.

[30] Questa astrazione viene usata nei linguaggi di descrizione dell'hardware (*Hardware Description Language*, HDL) per creare una rappresentazione di alto livello di un circuito digitale. Da tale rappresentazione vengono derivati i componenti di basso livello (porte, flip-flop, ..) e le loro interconnessioni.

Esse si appoggiano sì sul Sistema Operativo, ma anche sulle cosiddette API (*Application Programming Interface*), che, a loro volta, si appoggiano sul Sistema Operativo. In sostanza le applicazioni si appoggiano almeno su due livelli.

1.6 Siti web

Il modo più ovvio per cercare dati su un certo argomento è fare ricorso a un motore di ricerca, digitando, per esempio "Eniac".

Dati e curiosità sugli argomenti trattati da questo capitolo sono ampiamente disponibili su una moltitudine di siti web. La storia del calcolo, presente nella *Stanford Encyclopedia of Philosophy*, si trova, in forma più approfondita di quella tratteggiata al Paragrafo 1.1.1, all'indirizzo `http://plato.stanford.edu/entries/computing-history/`.

Sul sito del Museo Nazionale della Scienza e della Tecnologia "Leonardo da Vinci" di Milano, e in particolare all'indirizzo
`http://www.museoscienza.org/approfondimenti/documenti/macchina_poleni/storia5.asp`
si trovano notizie interessanti circa le prime macchine di calcolo meccaniche, tra cui la Pascalina e la macchina aritmetica di Leibniz. Soprattuto si scopre che, ancor prima di Blaise Pascal, l'astronomo, geometra e matematico tedesco Wilhelm Schickard, nel 1623 aveva progettato una macchina calcolatrice, da lui denominata "Orologio da calcolo", di cui si sono salvati i disegni originali.

All'indirizzo `http://www.library.upenn.edu/exhibits/rbm/mauchly/jwmintro.html` si trovano dati curiosi su Eniac e dintorni.

Normalmente tutte le società hanno un sito con un indirizzo così strutturato: `http://www.nome.com`, dove `nome` è il nome della società. Di solito questi siti presentano una finestra per immettere una frase di ricerca ed è facile dirigersi velocemente verso l'argomento di interesse.

Internet è ormai l'unica sede dove si possono trovare manuali dettagliati. Fino alla prima metà degli anni '90, i manuali venivano ancora stampati in forma cartacea. All'epoca, i professionisti interessati corteggiavano i rappresentanti delle case produttrici per accaparrarseli appena usciti. Oggi le informazioni sono molto più a portata di click. Tanto per dare un'indicazione, i manuali Intel (Intel 64 and IA32 Architectures Software Developer's Manual) di fine 2016 superavano le 4.000 pagine PDF, molto dense.

La pagina `http://www.intel.com/pressroom/kits/quickreffam.htm` fornisce dati tecnici di tutte le CPU Intel, nelle differenti versioni, a partire dal micro 4004. Sul sito IBM cercare "mainframe history" per vedersi presentare una serie di rinvii a pagine contenenti notizie storiche e foto dei calcolatori prodotti dalla società.

La pagina `http://infocenter.arm.com/help/index.jsp` è il punto di accesso al vasto corpo di letteratura ARM. Con pochi passaggi è possibile trovare manuali, note tecniche e anche articoli di ricerca riguardanti il mondo dei processori ARM.

Sul sito `http://www.top500.org/` si trovano elencati i primi 500 supercomputer al mondo, in ordine di potenza decrescente. Vengono riportate le statistiche a partire dal 1993, due volte all'anno, a Giugno e a Novembre. I dati possono essere scaricati in forma di foglio elettronico. La Tabella 1.2 è stata ottenuta come sottoinsieme del foglio elettronico relativo a Novembre 2016. Il lettore è invitato a esaminare i contenuti del sito e verificare come la potenza di calcolo è andata crescendo negli anni.

All'indirizzo `http://www.netlib.org/benchmark/performance.ps` si trova un articolo che illustra il benchmark Linpack e le misure delle prestazioni per una messe di sistemi.

Alla pagina `http://www.nature.com/nature/journal/v406/n6799/full/4061047a0.html` della rivista *Nature* si trova l'articolo su limiti fisici e legge di Moore.

Un articolo sulla legge di Moore di R. Saracco può essere scaricato da questa pagina `http://www.apogeonline.com/libri/88-503-1063-3/scheda`.

Domande ed esercizi

1.1 Si discuta il concetto di architettura e organizzazione. In particolare si approfondisca il concetto di astrazione e si discuta come questo viene applicato sia nella vita quotidiana sia in ambiti tecnologici.

1.2 Si ponga pari a 1 la velocità (le prestazioni) di una CPU nel 1996. Si assuma che il tasso di crescita per la velocità delle CPU dal 1995 ad oggi sia stato del 45%, con una crescita del 30% in conseguenza del miglioramento delle tecnologie di fabbricazione e una crescita del 15% in conseguenza delle soluzioni architetturali. Quali sarebbero le prestazioni alla data odierna se la crescita fosse stata solo tecnologica? Quali sarebbero le prestazioni se la crescita fosse stata solo architetturale?

1.3 Si assuma che nel 1986 una Ferrari percorresse il circuito di Monza in 1,5 minuti. Quanto impiegherebbe oggi se il tasso medio di crescita della velocità delle automobili fosse stato pari a quello delle CPU assunto pari a 45% all'anno?
Naturalmente nel risolvere l'esercizio non si deve tener conto delle limitazioni di velocità imposte dalle curve!

1.4 Si supponga che le prestazioni di una linea di calcolatori siano determinate al 50% dalla CPU e per il restante 50% da memoria e dischi. Poste pari a 1 le prestazioni nel 1995, di quanto sarebbero aumentate complessivamente all'anno 2000, assumendo i tassi di crescita delle rispettive velocita del 60% per le CPU e del 7% per i dischi?
Si ripeta lo stesso esercizio assumendo che la CPU abbia un impatto del 30% e la memoria e i dischi del 70%.

1.5 Un produttore di calcolatori costruiva nel 2007 macchine con prestazioni pari a 100. Nel 2016 le prestazioni dei suoi sistemi sono pari a 200. Come si situa questa produzione rispetto allo sviluppo tecnologico, assumendo che le prestazioni delle CPU siano mediamente cresciute del 30% all'anno, quelle della memoria del 10% e quelle dei dischi del 7%?
Si discuta la questione tenendo conto del fatto che le prestazioni non sono determinate dalla sola CPU (vedi Esercizio 1.4).